

## PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2001-005485

(43)Date of publication of application : 12.01.2001

(51)Int.Cl.

G10L 15/18

G06F 3/00

G06F 3/16

G10L 15/00

G10L 15/28

G10L 15/22

(21)Application number : 2000-149693

(71)Applicant : INF STORAGE DEVICES INC

(22)Date of filing : 22.05.2000

(72)Inventor : GEILHUF MICHAEL  
MACMILLAN DAVID  
BAREL AVRAHAM  
BROWN AMOS  
BOOTSMA KARIN LISSETTE  
GADDY LAWRENCE KENT  
PYO PHILLIP PAUL

(30)Priority

Priority number : 99 316604

Priority date : 21.05.1999

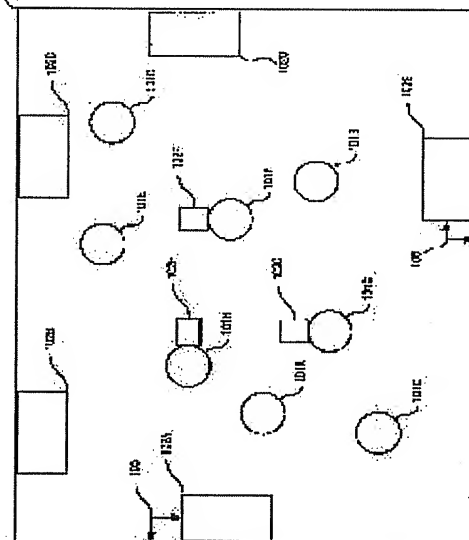
Priority country : US

## (54) METHOD AND DEVICE TO IMPROVE ACTIVITY OF VOICE CONTROL DEVICE

(57)Abstract:

PROBLEM TO BE SOLVED: To improve recognition precision by activating a voice control electronic device when all of the relative silence, name of the device and a command are recognized by the device.

SOLUTION: Voice control type devices 102A to 102H having an uttering recognition system, which waits for relative silence in a communication environment, are provided. Assume that human users 101A to 101H are conducting a conference in an environment 100, i.e., a conference room. Then, telephones having a speaker/phone function are assumed to be device names, i.e., voice control type devices 102A to 102H. The users 101A to 101H call for the name, i.e., a telephone and commands are given to the devices 102A to 102H. By giving the name of the devices 102A to 102H, the devices 102A to 102H are correctly responded to the given commands, and no confusion occurs between many users and the devices 102A to 102H.





(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号  
特開2001-5485  
(P2001-5485A)

(43) 公開日 平成13年1月12日 (2001.1.12)

(51) Int.Cl. <sup>7</sup>	識別記号	F I	テマコード* (参考)
G 1 0 L 15/18		G 1 0 L 3/00	5 3 7 G
G 0 6 F 3/00	6 0 1	G 0 6 F 3/00	6 0 1
	3 2 0		3 2 0 H
G 1 0 L 15/00		G 1 0 L 3/00	5 5 1 P
15/28			5 6 1 D

審査請求 未請求 請求項の数 8 O L (全 57 頁) 最終頁に続く

(21) 出願番号 特願2000-149693 (P2000-149693)

(22) 出願日 平成12年5月22日 (2000.5.22)

(31) 優先権主張番号 0 9 / 3 1 6 6 0 4

(32) 優先日 平成11年5月21日 (1999.5.21)

(33) 優先権主張国 米国 (U S)

(71) 出願人 595023426

インフォメーション・ストレージ・デバイ  
ーズ・インコーポレーテッド

アメリカ合衆国 95125 カリフォルニア  
州 サン ホゼ・ノース ファースト ス  
トリート・2727

(72) 発明者 マイケル・ゲイルフーフ

アメリカ合衆国・94301・カリフォルニア  
州・バロ アルト・テニソン アベニュー・  
619

(74) 代理人 100064621

弁理士 山川 政樹

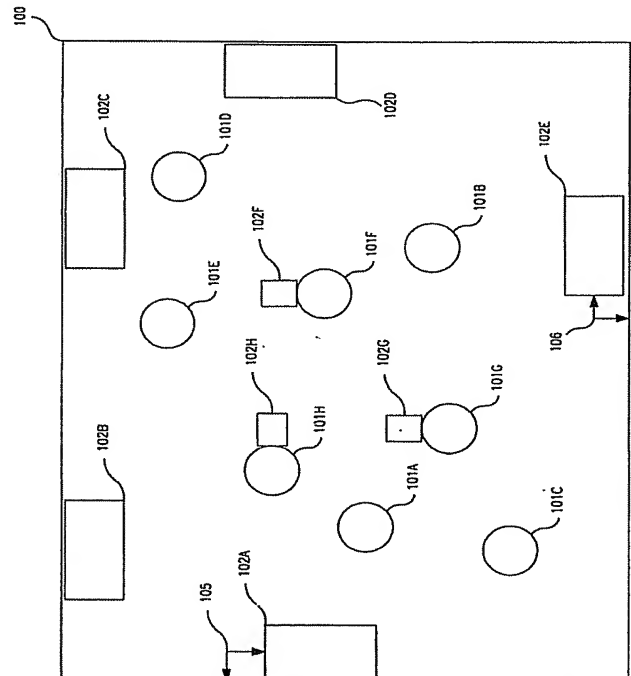
最終頁に続く

(54) 【発明の名称】 音声制御デバイスの活動化を向上させる方法および装置

(57) 【要約】

【課題】 認識精度を向上させるために音声制御電子デ  
バイスを活動化させる方法を提供する。

【解決手段】 この方法は、通信環境における相対的無  
音の期間を設けること、その音声制御デバイスに関連付  
けられた名前およびコマンドを伝達することを含む。音  
声認識エンジンによっていずれも認識されない場合、音  
声制御デバイスは活動化されない。音声制御デバイ  
スは、通信環境における相対的無音の期間を認識するこ  
とに従って活動化させるソフトウェア、前記音声制御デバ  
イスに関連付けられた名前およびコマンドを通知するこ  
とを含む。



## 【特許請求の範囲】

【請求項 1】 認識精度を向上させるために音声制御電子デバイスを活動化させる方法であって、

- a) 通信環境における相対的沈黙を待つ発話認識システムを有する音声制御電子デバイスを設けるステップと、
- b) 前記通信環境における第 1 の相対的沈黙の期間を設けるステップと、
- c) 前記第 1 の相対的沈黙の期間を設けた後に、前記通信環境で前記音声制御電子デバイスに関連付けられた器具名を通信するステップと、
- d) 前記通信環境でコマンド・セットのコマンドを通信するステップと、
- e) 前記相対的沈黙、前記器具名および前記コマンドがすべて前記音声制御電子デバイスによって認識された場合に前記音声制御電子デバイスを活動化させ、そうでない場合は認識されるまでステップ (b) から (e) のシーケンスを繰り返すステップとを含む方法。

【請求項 2】 前記音声制御電子デバイスが話者に依存しない請求項 1 に記載の音声制御電子デバイスを活動化させる方法。

【請求項 3】 認識精度が向上された音声制御電子デバイスであって、  
プロセッサと、  
前記プロセッサに結合されたプロセッサ可読記憶媒体と、  
前記プロセッサによって実行可能な前記プロセッサ可読記憶媒体に記録され、通信環境における第 1 の相対的沈黙の期間を認識するコードと、  
前記プロセッサ可読記憶媒体に記録され、通信環境において前記音声制御電子デバイスに通信された前記音声制御電子デバイスに関連付けられた器具名を認識するコードと、  
前記プロセッサ可読記憶媒体に記録され、前記プロセッサ可読記憶媒体に格納されたコマンド・セットのコマンドを認識するコードであって、前記コマンドが前記通信環境において前記音声制御電子デバイスに通信されるコードと、  
前記プロセッサ可読記憶媒体に記録され、前記第 1 の相対的沈黙の期間、前記器具名および前記コマンドを認識することに応答して前記音声制御電子デバイスを制御するコードとを含む音声制御電子デバイス。

【請求項 4】 前記プロセッサ可読記憶媒体に記録され、第 2 の期間内で前記音声制御電子デバイスに通信された前記コマンドに関連付けられたフレーズを認識することに応答して、前記音声制御電子デバイスをさらに制御するコードをさらに含む請求項 3 に記載の音声制御電子デバイス。

【請求項 5】 可聴音声を使用して前記器具名および前記コマンドが与えられる請求項 3 に記載の音声制御電子デバイス。

【請求項 6】 非可聴音声を使用して前記器具名および前記コマンドが与えら得る請求項 3 に記載の音声制御電子デバイス。

【請求項 7】 認識精度を向上させるために音声制御電子デバイスを活動化させる方法であって、

- a) 通信環境における相対的沈黙を待つ音声認識システムを有する音声制御電子デバイスを設けるステップと、
- b) 前記通信環境における第 1 の相対的沈黙の期間を設けるステップと、
- c) 前記第 1 の相対的沈黙の期間を設けた後に、前記通信環境で前記音声制御電子デバイスに関連付けられた器具名を通信するステップと、
- d) 前記通信環境でコマンド・セットのコマンドを通信するステップと、
- e) 前記第 1 の相対的沈黙の期間、前記器具名および前記コマンドを認識し、第 2 の期間の時間待つか、あるいは前記第 2 の期間の時間が満了する前に前記コマンドに関連付けられたフレーズが通信されるまで待つて、前記音声制御電子デバイスを活動化させ、そうでない場合は認識されるまでステップ (b) から (e) のシーケンスを繰り返すステップとを含む方法。

【請求項 8】 前記第 1 の相対的沈黙の期間、前記器具名および前記コマンドが認識され、前記フレーズが認識されない場合、前記音声制御電子デバイスが、前記フレーズが認識された場合とは異なる方法で応答する請求項 7 に記載の音声制御電子デバイスを活動化させる方法。

## 【発明の詳細な説明】

## 【0001】 マイクロフィッシュの付録

本出願は、ここには印刷されていないが参照により本明細書に組み込まれ、ここに完全に記述されているかのように逐語的であり同じ効力を有する、Information Storage Devices Inc. による「ISD-SR 300, Embedded Speech Recognition Processor」という名称のマイクロフィッシュの付録を含む。

## 【0002】

【発明の属する技術分野】 本発明は一般にマシン・インタフェースに関する。特に、本発明はデバイス用の音声ユーザ・インタフェースに関する。

## 【0003】

【従来の技術】 コンピュータ用のグラフィカル・ユーザ・インタフェース (GUI) はよく知られている。GUI は人間とコンピュータの対話のために、直感的で一貫した方法を提供する。一般に、一度特定の GUI の使用方法を学ぶと、学んだ人は同じまたは同様の GUI を使用して動作する任意のコンピュータまたは装置を動作できるようにする。普及している GUI の例は、Apple による MAC OS、Microsoft による MS Windows (登録商標) である。GUI は現在他の装置にも移植されている。たとえば、MS Wind

ows GUIは、いくつかの異なる装置の間に共通の GUIを与えるために、コンピュータからパーム・トップ、パーソナル・オーガナイザ、および他の装置へ移植されている。しかし、名前が示すようにGUIは少なくとも何らかの種類のビジュアルまたはグラフィカル・ディスプレイと、キーボード、マウス、タッチ・パッド、またはタッチ・スクリーンなどの入力デバイスを要求する。ディスプレイと入力デバイスは装置内のスペースを使用し、追加の構成要素を要求し、装置のコストを増大する傾向がある。したがって、装置からディスプレイと入力デバイスを除去し、コストを節約することが望ましい。

【0004】最近、装置を制御するために発話認識方法を使用する音声ユーザ・インタフェース（VUI）が導入された。しかし、これらの従来技術のVUIはいくつかの欠点を有し、これらがすべての装置で普遍的に使用できなくなっている。従来技術のVUIは通常使用が難しい。従来技術のVUIは通常、LCDなどの何らかのディスプレイを要求するか、キーパッドまたはボタンなどの手動の入力デバイスを要求するか、またはディスプレイと手動の入力デバイスの両方を要求する。さらに、従来技術のVUIは通常、専用であり、ハードウェア装置の単一の製作またはモデル、または単一のタイプのソフトウェア・アプリケーションへの使用に制限される。これらは通常、コンピュータのオペレーティング・システムとは異なって幅広く使用可能なわけではなく、したがって、ソフトウェアのプログラマは種々の装置内でVUIで動作するアプリケーションを書くことができない。従来技術のVUIに関連するコマンドは、通常は単一のタイプの装置またはソフトウェア・アプリケーションのためにカスタマイズされている。従来技術のVUIは通常、個別化とセキュリティを扱う方法など、多数のユーザをサポートすることにおいて追加の制限を有する。さらに従来技術のVUIは、前もってユーザがVUI用の装置の存在を知っていることを要求する。従来技術のVUIはその装置の存在を決定する方法を提供していない。さらに、従来技術のVUIは通常、インストラクション・マニュアルまたは画面に表示されたコマンドを読んでその使用に習熟することをユーザに要求する。従来技術のVUIは通常、ユーザがコマンドを学ぶための可聴方法を含まない。さらに、標準化が欠如しているため、多数の音声制御式デバイスを使用する時に多数の従来技術のVUIの使用方法を学ぶことを要求されることがある。

【0005】一般に、VUIによって制御された装置は、何らかの種類の手動制御を依然として要求する。何らかの手動制御が要求されているため、ボタン、キーパッド、または1組のボタンまたはキーパッドなどの手動入力デバイスが用意されている。正しく手動入力するために、LCD、LEDまたは他のグラフィックスディス

プレイなどのディスプレイが提供されることがある。たとえば、多くの音声起動電話は、電話番号が手動で格納されていることを要求する。この場合は通常、手動入力のために数字のキーパッドが用意されている。通常、正しく手動入力し、装置の状態を表示するためにLCDが含まれている。発話合成システムまたは音声フィードバック・システムは、これらの装置には欠けていることがある。ボタンとディスプレイの追加は、装置の製造コストを増大する。コストを下げるために、装置からすべての手動入力と表示を除去できることが望ましい。さらに、特定のボタンまたはディスプレイを必要とせずに、遠隔から装置を制御できるとさらに好都合である。

【0006】従来は、VUI用の装置は少数の人によって使用されていた。さらに、彼らは音声をローカルに聞くために近距離音場マイクロフォンを使用していた。多くの従来の装置は何らかの方法で固定されていたかまたは、携帯可能になりにくかったか、サーバに基づいたシステムであった。携帯装置のための音声制御機能を提供することが望ましい。音声制御式デバイスに、近距離音場マイクロフォン技術または遠距離音場マイクロフォン技術のどちらかを提供することが望ましい。音声制御機能がより多くの装置に含まれるように、低コストの音声制御機能を提供することが望ましい。しかし、これらの希望は多数の音声制御式デバイスおよび多数のユーザが同じ地域にいた時に問題を起す。たがいに可聴範囲内に多数のユーザと多数の音声制御式デバイスがあると、音声制御式デバイスがどのユーザからのコマンドを受け取り、応答するのか見分けることが難しくなる。たとえば、多数のユーザがいる環境内で1人のユーザが家に電話をかけたい場合の、音声制御セル電話の場合を考えてみる。ユーザは音声で起動された、家に電話をかけるコマンドを発行する。複数の音声制御セル電話が家に電話をかけるコマンドを可聴に聞いた場合、多数の音声制御セル電話が応答し、それぞれの家の電話番号のダイヤルを開始する。以前は、音声制御式デバイスがほとんどなかったのもこれはそれほど大きな問題ではなかった。

【0007】

【発明が解決しようとする課題】いくつかの音声制御式デバイスは発話者に依存する。発話者依存性とは、特定のユーザが使用できるようになる前に、そのユーザによる訓練を要求する音声制御式デバイスを指す。発話者に依存した音声制御式デバイスは、フレーズが話される方法における音色の質を聞く。発話者に依存した音声制御式デバイスは、多数のユーザまたは発話者が音声制御式デバイスを使用するように要求されている用途には向いていない。これは、これらの装置は訓練を受けていないユーザからの発話を効率的に認識することができないためである。任意のユーザからの発話を認識するために、わずかな訓練を要求するかまったく訓練を要求しないVUIを備えた発話者から独立した音声制御式デバイスを

提供することが望ましい。

【0008】高い精度の発話認識を達成するために、音声制御式デバイスはその装置に向けられていない発話への応答を回避することが重要である。すなわち、音声制御式デバイスは背景の会話、雑音、他の音声制御式デバイスへのコマンドに応答すべきではない。しかし、背景の音を濾過して取り除くことが効果的すぎて、音声制御式デバイスに向けられた発話の認識を妨げてはならない。背景の音の拒否と、音声制御式デバイスへ向けられた発話の認識の正しい混合比を見つけることは、発話者から独立したシステムには特に難しい。発話者から独立したシステムでは、音声制御式デバイスは幅広い声に応答できなければならない、したがって、背景の音について制限の厳しいフィルタを使用することはできない。対照的に、発話者に依存したシステムは特定の人の声だけを聞く必要があり、したがって背景の音に関してさらに厳重なフィルタを使用することができる。発話者に依存したシステムにおけるこの利点にもかかわらず、背景の音を濾過して取り除くことは依然として大きな課題である。

【0009】いくつかの従来技術のシステムでは、発話の認識を起動するためにユーザにボタンを物理的に押させることによって背景の会話が濾過して取り除かれた。この方法の欠点は、ユーザが音声または発話だけによってではなく、音声制御式デバイスと物理的に対話することを要求することである。音声制御式デバイスの潜在的な利点の1つは、音声制御式デバイスがまったくハンドフリーの動作を約束することである。発話の認識を起動するためにボタンを押す必要性を除去することは、このハンドフリーの目的を達成するために多いに役立つであろう。

【0010】さらに、何人かの人が話している場所では、音声制御式デバイスは発話がその装置に向けられていない限り、すべての発話を見逃すべきである。たとえば、1人の人が別の人に「私はジョンに電話をする」と言った場合、彼のポケットにあるセル電話は「ジョンに電話をします」をコマンドとして翻訳すべきではない。1つの場所に多数の音声制御式デバイスがある場合、ユーザが制御したいのはどの音声制御式デバイスであるかを一意的に識別する方法があるべきである。たとえば、多数の音声制御電話、おそらく1組のデスクトップ電話と、多数のセル電話がそれぞれ1人に1つずつある部屋を考えてみる。だれかが「555-1212に電話をします」と言ったとすると、各電話について所定のコマンドを見逃すための手段がない限り、各電話は電話をかけようと試みることがある。1つの音声制御式デバイスが多数のユーザによって制御される場合、音声制御式デバイスがどのユーザがそれにコマンドを与えているのか知ることが望ましい。たとえば、家の中の音声制御デスクトップ電話は、夫、妻、子供によって使用されること

がある。各人は頻繁に電話する番号の独自の電話番号を有することがある。音声制御式デバイスが「母に電話します」と言われた時、装置が正しい人に電話できるように、どのユーザがコマンドを発行しているのか知る必要がある（つまり、電話は夫の母に電話すべきなのか、妻の母に電話すべきなのか、または子供の母の仕事用の番号に電話すべきなのか）。さらに、多数のユーザを有する音声制御式デバイスは、許可されない使用から守るためのセキュリティを実行したり、またはユーザの個別化を他の人による故意でない対話または悪意のある対話（盗み聞き、設定の変更、削除、追加を含む）から守るための方法を必要とすることがある。さらに、多数の音声制御式デバイスがある場所では、音声制御式デバイスの存在を識別する方法があるべきである。たとえば、新しいホテルの部屋についた旅行者を考えてみる。ホテルの部屋に入った時、旅行者はどんな音声制御式デバイスが存在するのか、およびそれらをどのように制御するのかを知りたいと思う。すべての音声制御式デバイスが同じ方法で識別できるように、識別プロセスが標準化されていることが望ましい。

【0011】音声制御式デバイスでは音声制御のもとでフレーズを格納することが望ましい。フレーズは単一の言葉、または単位として扱われる1グループの言葉として定義される。この格納は、選択肢を設定することまたは個別化を作成することであることがある。たとえば音声制御電話では、音声制御の元で人の名前および電話番号を個別化された電話帳に格納することが望ましい。後から、この電話帳を使用して人の名前を発話することによって彼らに電話することが可能である（たとえば、「セル電話、ジョン・スミスに電話します」または「セル電話、母に電話します」）。

【0012】フレーズ（「ジョン・スミス」）を格納するための従来技術の方法は、実際の音を保存しようとする圧縮された方法、圧縮されない方法、または変形された方法でフレーズを格納することによって動作した。次いで、コマンド内におけるフレーズの検出は（つまり、上の例ではジョンが電話されるべきであることを検出すること）、元の格納された発話音と、話されたコマンドの間の、音に基づいた比較に依存する。時々、格納された波形は周波数ドメインに変形されるか、および/または、一致を促進するために時間調節されるが、どの場合でも実行される基本的な動作は、実際の音を比較することである。格納された音の表現と検出のための比較は、いくつかの欠点を被る。おそらく寒さ、ストレス、疲れ、電話による騒音またはゆがんだ接続、または他の要因によって発話者の声が変わると、典型的に比較は成功せず、格納されたフレーズは認識されない。フレーズは音の表現として格納されているので、テキストに基づいたフレーズの表現を抽出する方法はない。さらに、音表現を格納する結果、発話者に依存したシステムとなる。

別の人がコマンドで同じ音を使用して同じフレーズを話し、フレーズを正しく認識させることが可能であるとは考えられない。たとえば、秘書が電話帳のエントリを格納し、マネジャがこれらのエントリを使用して電話をすることは信頼性がなくなる。発話者から独立した格納手段を提供することが望ましい。さらに、フレーズが音表現として格納されている場合、両方の音声制御式デバイスによって同じ波形処理アルゴリズムが使用されていない限り、格納されたフレーズは別の音声制御式デバイス内で使用できない。フレーズは一度格納されたら、そのフレーズを使用して発話者から独立して認識でき、多数の音声制御式デバイスによって使用できるような表現で話されたフレーズを認識し格納することが望ましい。

【0013】現在、コンピュータと他の装置はモデム、赤外線、または無線高周波伝送を使用して他のコンピュータや装置にコマンドおよびデータを通信している。送信されたコマンドおよび／またはデータは、通常はそのコンピュータまたは装置だけが理解できるデジタルな形である。人間のユーザがコマンドまたはデータを理解できるようにするためには、コマンドまたはデータはコンピュータによって復号化され、次いでディスプレイ上に数字またはASCIIテキストなどの何らかの種類のフォーマットで表示されなければならない。コマンドおよび／またはデータが送信される時、これらは通常はコンピュータまたは装置または送信機器によって理解される何らかのデジタルなフォーマットにコード化される。音声制御式デバイスがより普及するにつれて、音声制御式デバイス間の通信のための追加の回路を設けるのを避けるために、音声制御式デバイスが人間のような発話を使用して互いに通信できることが望ましい。さらに、人間のユーザの介入がなくても、多数の音声制御式デバイスがマシンからマシンへ情報を交換できることが望ましい。

#### 【0014】

【課題を解決するための手段】本発明は請求項に説明されたような方法、装置、およびシステムを含む。簡単に言うと、標準の音声ユーザ・インタフェースが提供され、標準の発話コマンドを使用することによって種々の装置を制御する。標準のVUIはユーザと音声制御式デバイス間のインタフェースに関し、1組の標準のVUIコマンドと構文を提供する。標準のVUIコマンドは、音声制御式デバイスが1つの環境の中で使用可能かどうかを決定するための識別フレーズを含む。他の標準VUIコマンドは、音声制御式デバイスの名前を決定し、これらを変えるために設けられる。

【0015】音声制御式デバイスが開示される。音声制御式デバイスは、ここでは可聴発話または非可聴発話のどちらかによって制御される任意の装置として定義される。音声制御式デバイスはまた、ここでは器具、マシン、音声制御器具、音声制御電子器具、名前起動電子器

具、発話制御装置、音声起動電子器具、音声起動器具、音声制御電子装置、または自動識別音声制御電子装置とも呼ばれることがある。

【0016】音声制御式デバイスの機能へアクセスするために、ユーザは一定期間の相対的な沈黙の後に、関連する器具名のうち1つを音声制御式デバイスに通知する。器具名はデフォルトの名前の可能性もあり、ユーザが割当て可能な名前の場合もある。音声制御式デバイスは、装置に関連した、ユーザが割当て可能な複数の名前を有し、各ユーザに対して個別化された機能を提供する可能性もある。

【0017】本発明の他の態様は、詳細な説明の中で説明される。

#### 【0018】

【発明の実施の形態】以下の、本発明の詳細な説明の中では、本発明の完全な理解を与えるために多くの特定の詳細が設定されている。しかし当業者であれば、これらの特定の詳細がなくても本発明が実行できることは明らかであろう。他の例では本発明の態様を不必要に曖昧にしないために、良く知られた方法、手順、構成要素、回路は詳細には説明されていない。

【0019】本発明は、標準の音声ユーザ・インタフェースと音声制御式デバイスのための方法、装置、およびシステムを含む。簡単に言えば、標準の発話コマンドを使用することによって種々の装置を制御するために、標準の音声ユーザ・インタフェースが提供される。標準のVUIはユーザと音声制御式デバイス間のインタフェースに関して1組のコアVUIコマンドと構文を提供する。コアVUIコマンドは識別フレーズを含み、音声制御式デバイスがある環境の中で使用可能かどうかを決定する。他のコアVUIコマンドは、音声制御式デバイスの名前を決定し、名前を変更するためのものである。

【0020】音声制御式デバイスが開示される。音声制御式デバイスは、ここでは可聴発話または非可聴発話のどちらかによって制御される任意の装置として定義される。可聴および非可聴は、ここではのちに定義される。音声制御式デバイスはまた、ここでは器具、マシン、音声制御器具、音声制御電子器具、名前起動電子器具、発話制御装置、音声起動電子器具、音声起動器具、音声制御電子装置、または自動識別音声制御電子装置とも呼ばれることがある。

【0021】本発明は可聴発話および非可聴発話を使用して制御され、通信する。本発明に関してここで定義された発話は、a) 信号または情報がそれを空気圧の変化に変換する適切な装置を介して通過させられた場合、信号または情報は人間によって聞かれることができ言語として考えられるような信号または情報、b) 人間が信号を聞く場合は、人間がそれを言語として考えるように、空気圧の実際の変化を含む信号または情報を包含する。

可聴発話は、補助がなくても人間が聞くことのできる発話を指す。非可聴発話は、可聴発話の定義の元に含まれない発話の任意のコード化または表現を指し、人間の聴力範囲の外で、空気以外の送信媒体で通信される可能性のあるものを含む。発話の定義は、人間によって発せられた発話と、マシンから発せられた発話を含む（マシン発話合成、プロンプトまたは他の形式など前もって記録された人間の発話の再生を含む）。

【0022】音声制御式デバイスによって通知されるプロンプトと、ユーザによって通知されるフレーズは、英語以外の言語または方言、または多数の言語の組合せであることがある。フレーズはここでは単一の単語、または単位として扱われる単語のグループとして定義される。ユーザはここで定義されているように人間または装置であり、音声起動装置を含む。したがって「ユーザの話されたフレーズ」「コマンドを発行するユーザ」およびユーザによるすべての他の処置は、装置による処置と人間による処置を含む。

【0023】音声制御式デバイスは発話によって制御されるために、何らかの形の発話認識を含む。発話認識と音声認識はここでは同義語として使用されており、同じ意味を有する。好ましくは、発話者から独立した発話認識システムが、音声制御式デバイスの発話認識機能を提供するために使用される。発話者から独立した発話認識システムは、発話者から独立した発話表現に応答する。好ましい実施形態では、発話者から独立した発話表現は、発話の音声的な表現である。しかし、発話者から独立した発話の他の表現もまた、本発明によって使用されることがある。

【0024】本発明を備えた音声制御式デバイスの完全な機能へアクセスするために、ユーザは音声制御式デバイスが関連する器具名の1つを音声制御式デバイスに通知しなければならない。器具名は1つまたは複数のデフォルト名、または1つまたは複数のユーザが割当可能な名前を含む。音声制御式デバイスは、個別化された機能を各ユーザに提供するために、音声制御式デバイスに関連した、ユーザが割り当て可能な複数の名前を有することがある。

【0025】さらに本発明は、装置に向けられていない背景の騒音または発話によって間違えてトリガさせられることなく、会話が継続している間発話認識エンジンを起動したままにしておく方法を提供する（ローカルな会話または電話リンク上の会話を含む）。これを達成するために、本発明は本発明の標準VUIによって提供される音声制御式デバイス用の命名スキームを使用する。一般に、音声制御式デバイスはその器具名によって指名されない限り、装置はすべての発話を無視する。（のちに論じられるが、この規則には1組の特別な例外がある）。所定の場合では、コマンドを認識するための基準はさらに強化され、音声制御式デバイスのユーザが割り

当てた器具名によって指名されることを必要とする。音声制御式デバイスは多数のユーザを有する可能性があり、多数のユーザの各々は、本発明の標準VUIのコマンドを使用して装置に一意的な器具名を割り当てる。音声制御式デバイスがユーザ割当て名のうち1つによって指名された時、音声制御式デバイスはこれが指名されていることを決定し、またどのユーザが装置を指名しているのかを決定することができる。これは音声制御式デバイスがその、特定のユーザに関する個別化を使用することができる。たとえば、音声起動電話が4つの異なるユーザ割当て名を有し（たとえばAardvark, Platypus, Socrates, Zeus）、各ユーザは母に向けた電話帳エントリに関連した異なる電話番号を有することがある。第1のユーザが「Aardvark、母に電話をします」と言った時、第1のユーザの母が電話で呼ばれる。第2のユーザが「Platypus、母に電話をします」と言った時、第2のユーザの母が電話で呼ばれる。Geronimoはその器具名の1つではないので、「Geronimo、母に電話をします」というコマンドはこの音声制御式デバイスによって起動しない。

【0026】本発明の他の態様は、音声制御式デバイスの認識の精度を向上する。本発明は、第1に音声制御式デバイスに向けられたフレーズの前に相対的な沈黙の期間を要求し、第2に器具名を要求し、第3に有効なコマンドを要求することによって認識の精度を集合的に向上する。完全な沈黙は必要ないが相対的な沈黙が必要であり、ここで相対的な沈黙とは、フレーズが話されている時の音レベルより静かな音レベルと定義される。要求される特定の期間の相対的な沈黙と、相対的な沈黙と音声制御式デバイスに向けられた話されたフレーズの音強度との間の許容されたデシベル差は、作成される音声制御式デバイスのタイプ、装置が目的とする動作環境、使用される発話認識システムの能力、および他の要因に依存する。いくつかの場合では、要求される相対的な沈黙の長さおよび/またはデシベル差は、特定の環境内で認識の精度を最大化するために、音声制御式デバイスまたは関連する回路またはソフトウェアによって変化させられる可能性もある。標準VUIによれば、各ユーザは音声制御式デバイスに一意的な名前を割り当てることができ、またはデフォルトの器具名を使用することもできる。器具名を音声制御式デバイスに通知した後、コマンドが話されなければならない。この時点での有効なフレーズは、「Help」または「Cancel」のような特別なフレーズを含み、標準のVUI文法の一部である。有効なコマンドが認識されないと、音声制御式デバイスは全シーケンスを拒否し、沈黙を待つ状態に戻る。さらに、コマンドに応じて、典型的にはコマンドの修飾子を表す1つまたは複数の追加フレーズが提供されるまたは要求されることがある（たとえば、コマンド・シー



ケンス「<沈黙>555-1212に電話します」内の電話番号)。この時点での有効な入力または「Help」または「Cancel」のような特別なフレーズを含み、これは標準のVUI文法の一部である。コマンドの後短い時間内に有効なフレーズの検出に失敗することは、全コマンド・シーケンスを拒否するための基礎として使用するか、ユーザに彼の目的を明確にするようにプロンプト指示するための基礎として使用することができる。どちらの方法でも、これは、追加のレベルの精度チェックとして働く。別法として、フレーズがコマンドの後短い時間の間に検出されない場合でも、とにかくコマンドが実行される可能性もある。

【0027】音声制御式デバイスは、視覚的な識別、または音響的な識別、またはその両方によって識別できる。音響的な識別は、音声制御式デバイスとの可聴通信と非可聴通信の両方を含むとして定義される。可聴と非可聴は他の場所で定義されている。視覚的な識別は標準的なロゴまたは他の視覚的な識別子の使用を介して生じることがある。点滅するLEDは、視覚的な識別子の別の例である。視覚的な識別は特に、常に起動されている発話認識エンジンを有しない音声制御式デバイスに適している。たとえば、バッテリーの消費を最小にするために、バッテリーで作動する音声制御式デバイスはスイッチ（または、フリップ型セル電話のフリップ・オープンなど、その等価物）を押して発話認識エンジンを起動することをユーザに要求することがある。音響的な識別は、認識可能なコマンドを積極的に聞く音声制御式デバイスだけに機能する。音響的な識別は、識別フレーズを言うユーザによって達成される。識別フレーズの例は、「そこに何がありますか」である。音声制御式デバイスは、1つまたは複数の識別フレーズを有することがある。識別フレーズを聞く任意の音声制御式デバイスが応答して、装置の存在を識別する。標準VUIによれば、応答は最長2秒までの沈黙のランダム遅延であり、そのあとに標準の信号（たとえば1つまたは複数のトーンまたはビーという音または他の音）、次いで少なくとも1つの音声制御式デバイスの器具名、および任意の適用可能な基本的な動作命令が続く（たとえば、「<ビー>私は電話です。あなたは電話ヘルプとすることができます」）。同じ通信環境内で多数の音声制御式デバイスからの応答を調整するために、各音声制御式デバイスはその沈黙期間の間、別の音声制御式デバイスの応答を聞かなければならず、他の音声制御式デバイスの応答の開始は標準信号によってマークされる。他の音声制御式デバイスの標準信号の検出は都合のよい任意の手段で達成でき、音声認識システムによる手段、DSPによる手段、マイクロプロセッサによる手段、または特別な回路による手段を含む。別の音声制御式デバイスがこの沈黙期間の間に応答を開始した場合は、聞く側の音声制御式デバイスは、応答する側の音声制御式デバイスが終了した

後、沈黙タイミングを再開始しなければならない。2つの音声制御式デバイスがほとんど同時に応答を開始した場合（たとえば、それらがちょうど良い時に標準信号重複であるように）、これらは両方とも、ランダムに選択された新しい沈黙遅延の間譲歩しなければならないが、この時遅延は前の沈黙遅延の長さの最長2倍までであり、16秒を超えてはならない。

【0028】識別フレーズに応答する音声制御式デバイスを制限するために、ユーザは識別フレーズ内に音声制御式デバイスの名前を含めることがある。たとえば1人のユーザが「Socrates、そこにありますか」と言って、Socratesと名付けられた音声制御式デバイスが近くにあるかどうかを確かめることがある。同様に、1人のユーザが、「Clock、そこにありますか」と言って、Clockという器具名（デフォルトの器具名であろうとユーザ器具名であろうと）を有するすべての音声制御式デバイスに応答させる可能性もある。可能性のある変形例として、たとえばセキュリティ上の理由で必要とされるために、音声制御式デバイスがその名前以外の何らかの応答で応答することがある。

【0029】音声制御式デバイスは、視覚的な識別方法と音響的な識別方法の両方を使用することがある。たとえば、発話認識エンジンが連続的に動作中であっても、視覚的なロゴおよび/または他の視覚的な識別子を依然として表示することがある。同様に、発話エンジンの手動での起動を要求する音声制御式デバイス内では、エンジンは一度起動されると、「そこに何がありますか」というコマンドに応答できる。

【0030】本発明の別の態様では、ユーザの話されたフレーズの最初の格納（たとえば音声制御の元で新しい電話帳エントリを作成する時）は、音声制御式デバイスの、発話者から独立した発話認識エンジンによって処理される。このエンジンは、発話者から独立した、フレーズの音声的な表現を戻す。この発話者から独立した音声的な表現が、格納されたものである。

【0031】コマンドがユーザから発行された時、これはまた、本発明の発話者から独立した発話認識エンジンによっても処理される。これは元のエントリを格納するための、発話者から独立した同じエンジンの使用である可能性もあるし、まったく異なる、発話者から独立したエンジンである可能性もある。どちらの場合でもエンジンは、コマンド・シーケンスの発話者から独立した音声的な表現を戻す。この発話者から独立した音声的な表現は、前に格納された音声的な表現と比較され、コマンドが認識可能かどうかを決定できる。

【0032】格納されている話されたエントリと任意のコマンドの両方を、発話者から独立した音声的な表現に変換することによって、多くの利点が提供される。

・おそらく病気、ストレス、疲労、雑音の多いまたは歪んだ電話リンクでの送信、または人間のユーザまたはマ

シンのユーザの会話を変更する可能性のある他の要因によってユーザの音声に変化した場合でも、認識は信頼できる。テキストに基づいた情報が格納でき、次いで認識できる。

・認識は他のユーザが元の音声フレーズを格納した場合でも信頼できる。

・認識は、ユーザが格納したコマンドおよびフレーズについても、発話者から独立したものとなることがある。

・テキスト・ソースから発生し格納された入力および異なる発話者から発生し格納された入力がすべて組み合わせられ、信頼を持って認識できる。

・発話者から独立した音声上の表現の使用は、これらが使用可能になると、改良された認識エンジンへのアップグレードを促進する。改良された発話認識エンジンは、すべての格納されたエントリが音声的な形で保持されているため、信頼性に影響を与えることなくまたは再格納を要求することなく、既存の格納された情報を使用することができる。改良された発話認識エンジンを使用して格納された新しい情報は、より古い認識エンジンを有する機器上でも使用できる。古い世代の機器と新しい世代の機器は、音声的な表現を使用することによってあらかじめ調整しなくても対話できる。これによってたとえば、2つのPDAが音声で格納された電話帳エントリを交換し、その情報の新しいユーザに信頼できる認識を提供することが可能になる。最後に、同じレガシ波形の変換を常に正しく実行できなければならない波形格納に基づいたシステムとは異なり、発話者から独立した認識エンジンが音声的な表現を作成できる限り、発話者から独立した認識エンジンのさらなる開発を妨げたり制限したりするレガシの制限はない。

#### 【0033】音声制御式デバイス

次に図1Aを参照すると、環境100が示されている。環境100は、オフィス、会議室、ホテルの部屋のような通信環境、または音声制御式デバイスが位置する可能性のある任意の場所であることがある。環境100の中には、円によって表される何人かの人間のユーザ101A~101Hがいる。また環境100の中には、正方形および長方形によって表され、各々本発明の標準の音声ユーザ・インタフェース(VUI)によって動作するように制御される音声制御式デバイス102A~102Hがある。長方形によって表されている音声制御式デバイス102A~102Eは、環境100の中で固定されている。正方形によって表されている音声制御式デバイス102F~102Hは、人間のユーザ101F~101Hにそれぞれ関連づけられている移動音声制御式デバイスである。音声制御式デバイス102A~102Hはまた、既存の装置または将来の装置であることがある。音声制御式デバイス102A~102Eは一般に、ユーザの自動車、家、オフィス、工場、ホテル、または人間のユーザがいる可能性のある他の場所に関連づけられてい

ることがある。別法として、音声制御式デバイス102A~102Eが非可聴発話によって制御される場合、その音声制御式デバイスは任意の場所に位置できる。

【0034】本発明では、標準VUIを使用するとユーザは、ユーザが割当て可能な名前をこれらの音声制御式デバイス102A~102Hに関連付けることができる。音声制御式デバイスのユーザが割当て可能な名前は、電話、時計、光などの一般的な名前であることがある。別法として、名前はもともと人間に与えられるJohn、Jim、Georgeなどの人格化された名前である可能性もある。どちらの場合でも、音声制御式デバイス102A~102Hは連続的に聞いているが、その名前のうち1つ(ユーザ割当てまたはデフォルト)を認識するまでコマンドに応答しない。任意の名前を音声制御式デバイスに割り当てることができるが、音声制御式デバイスと実際の人の間の混乱を最小限にするために、ユーザは通常の会話の間には生じる可能性のないAardvarkまたはSocratesなどの普通でない名前を選ぶことがある。図1Aを参照して、環境100が人間のユーザ101A~101Hが会議をしている会議室であると考えてみる。さらに音声制御式デバイス102Aは、会議室100内でスピーカ・ホン機能を有する電話であり、器具名は電話であると仮定する。101Aなどの人間のユーザはまず、電話という名前を呼び出してから、その音声制御措置にコマンドを与える。音声制御式デバイスに名前を与えることにより、音声制御式デバイスは与えられたコマンドに正しく応答し、多数のユーザと音声制御式デバイスの間の混同を避けることができる。音声制御式デバイスは電話、オーガナイザ、計算機、照明、ステレオ・システム、電子レンジ、TVセット、洗濯機、ドライヤ、暖房システム、冷房システム、または実質的に任意のシステムとすることができる。音声制御式デバイス102A~102Hは、人間のユーザ101A~101Hからのコマンドとデータ入力を聞き、コマンドまたはデータが正しく解釈され実行されていることをユーザに可聴的に知らせるために、可聴通信インタフェース(ACI)を含むことがある。音声制御式デバイス102A~102Hはさらに、発話認識及び合成システム(SRS)を含む。SRSの発話認識は、発話しているユーザからは独立して、およびユーザが人間であるか装置であるかを問わず、異なる言語の発話の解釈を行う。本発明の好ましい実施形態は発話者から独立した音声認識システムを使用しているが、本発明はまた、発話者に依存した音声認識システムとも互換性がある。SRSは、1つまたは1つ以上の言語で動作できる。SRSの発話合成は、可聴的に通信される可能性もありまたは非可聴的に通信される可能性もある音声制御式デバイスによる発話応答、状態コマンド、またはデータを生成する。ここではまた発話の生成とも呼ばれる発話合成は、ここで定義され、発話と応答する任意の方法

を含み（可聴または非可聴）、発話の記録、格納と再生システム、再生を伴う以前に記録されたボキャブラリ・システム、文字の組合せから発声を生成する洗練された発話合成システム、および上記の何らかの組合せを含むが、これらに限定されるものではない。好ましくは、音声制御式デバイスは発話の記録、格納と再生システム、および再生を伴う、以前に記録されたボキャブラリ・システムの両方を含む。

【0035】音声制御式デバイス102A~102Hはオプションとして、通信インタフェース（ECI）を含み、非可聴音声または会話を使用した無線手段または有線手段を介して音声制御式デバイスの遠隔制御を行う。図1Aに示されたように音声制御式デバイス102Aは、電話システムへの接続のために接続105を有する。この方法では、音声制御式デバイス102Aは遠隔からユーザに通知し、コマンドを受け取り、認識する。次に図1Bを参照すると、人間のユーザ101Iは、電話会社のスイッチ116上で、有線伝送媒体114または無線伝送媒体114上で電話112によって通信する。電話会社のスイッチ116は、接続105を介して有線手段または無線手段によって音声制御式デバイス102Aに接続されている。電話112は無線電話または有線電話である。この方法で、人間のユーザ101Iは、遠隔から通信環境100内の音声制御式デバイス102Aへ遠隔からインタフェースすることがある。別法として、音声制御式デバイス102Eなどの音声制御式デバイスは、遠隔コンピュータ118によってネットワーク上で遠隔から制御されることがある。この場合、遠隔の人間のユーザ101Jは、ネットワーク接続120と接続106を介して音声制御式デバイス102Eに結合されている遠隔コンピュータ118を介して、音声コマンドまたは命令を送信することができる。ネットワーク接続120は、インターネットなどのコンピュータ・ネットワークを介した、リアルタイム送達または蓄積交換の無線接続または有線接続であることがある。遠隔ユーザが音声制御式デバイスに接続できる幅広い方法があり、その中には有線接続および無線接続の使用が含まれるが、これに限定されるものではない。有線接続は電話システムおよびリアルタイムのインターネット接続などのリアルタイム通信システム、音声表現の電子メールおよび他の非リアルタイムのインターネット・プロトコルなどの蓄積交換システムを含むことがあるが、これに限定されるものではない。無線システムは無線システムおよび赤外線システムを含むことがあるが、これに限定されるものではない。これらの任意の代替例は、回路に基づいたシステムおよびパケットに基づいたシステムを含む可能性があり、アナログ・システムおよびデジタル・システムを含むことがある。これらの任意の代替例は、種々の変調スキームおよび／またはコード化スキームおよび／または暗号化スキームと共に使用される可能

性もあり、またはこれらなしでも使用されることがある。

【0036】次に図2を参照すると、例としての音声制御式デバイス102I~102Mが図示されている。音声制御式デバイス102Iは、冷凍庫、冷蔵庫、洗濯機、ドライヤ、エア・コンディショナ、暖房装置、電子レンジ、オーブン、およびストーブなどの家電製品の例である。音声制御式デバイス102Jは、オプションの通信インタフェース（ECI）を要求する音声制御式デバイスの例である。これはテレビ、ビデオ・カセット・レコーダ、ステレオ、カムコーダ、テープ・レコーダ、口述装置、目覚まし時計、時計ラジオ、および、標準の有線電話、電話応答マシンなどの電話製品、照明スイッチ、警報システム、計算装置、インターネット接続装置、サーバなどの、消費者家電のための音声制御式デバイスを含むことがある。音声制御式デバイス102Kは、セルラ電話、ウォークマン（登録商標）・スタイルのシステム、カムコーダ、パーソナル・デジタル・システムなどの携帯システムまたは無線システムの例である。音声制御式デバイス102Lは、自動車セルラ電話システム、自動車無線システム、カー・ナビゲーション・システム、HAV（暖房、エア・コンディショニング、換気）システム、および自動車用の他の制御システムなどの、自動車音声制御システムの例である。音声制御式デバイス102Mは、ボイスメール・システムなどの遠隔制御装置の例である。

【0037】音声制御式デバイス102Iは、可聴通信インタフェース（ACI）202、発話認識および合成システム（SRS）204、および電気製品周辺機器と制御回路（APCC）206を含む。ACI202はSRS204に結合され、SRS204はAPCC206に結合されている。音声制御式デバイス102Iでは、ACI202は発話通信の主な手段である。

【0038】音声制御式デバイス102JはACI202、SRS204、APCC206、通信インタフェース（ECI）207、および接続208を含む。ACI202はSRS204に結合されている。APCC206はSRS204に結合されている。ECI207はSRS204に結合し、接続208はECI207に結合している。音声制御式デバイス102Jは代替として、ACI202またはECI207を介して発話通信信号または音声通信信号を使用して通信できる。音声制御式デバイス102Kは、ACI202、SRS204、APCC206およびアンテナ209を含む。

【0039】音声制御式デバイス102Kは、ACI202を介した可聴発話信号を使用して、またはECI207を介したコード化された発話信号を使用して通信できる。ECI207はAPCC206に結合している。ECI207はまた、接続212に結合している。接続212はたとえば、アンテナまたは赤外線ポートである

ことがある。音声制御式デバイス102Lはまた、ACI202、SRS204、APCC206およびアンテナ209を含む。ACI202はSRS204に結合する。SRS204はAPCC206に結合する。アンテナ209はAPCC206に結合する。音声制御式デバイス102Lは、アンテナ209を介してACI202およびAPCC206によって通信できる。

【0040】音声制御式デバイス102Mは、APCC206、SRS204、ECI207、および接続210を含む。接続210は有線接続または無線接続である可能性があり、アンテナを含む。SRS204はAPCC206に結合し、またECI207にも結合する。接続210はECI207に結合する。音声制御式デバイス102Mは、接続210上で、ECI207を介して通信できる。

【0041】APCC206は、制御されるべき音声制御式デバイス102の要素を表す。たとえば家電製品の場合、制御されるべきアイテムは用途に応じて、温度、時間設定、電力設定、またはサイクルであることがある。消費者用電子製品の場合は、APCC206は通常ボタン、スイッチ、ノブに関連するアイテムからなりたつことがある。電話製品の場合、APCC206はボタン、ダイヤル、ディスプレイ、有線呼または無線呼を行うための回路または無線装置を表すことがある。自動車システムの場合、APCC206は、計測器パネル、温度ノブ、ナビゲーション用システム、自動車無線チャネル、音量、周波数特性を表すことがある。

【0042】次に図3を参照すると、音声制御式デバイス102が図示されている。図3に示された音声制御式デバイス102は、ここに説明された音声制御式デバイス内の機能ブロックの例である。音声制御式デバイス102はACI202、APCC206およびSRS204を含む。音声制御式デバイス102はまた、ECI207AまたはECI207BなどのECI207を有することができる。

【0043】図3に示されたACI202は、マイクロフォン303、スピーカ304、増幅器305を含む。図3に示されたようなSRS204は、音声通信チップ301、コード器／復号器（コーデック）306および308、ホスト・マイクロコントローラ310、電源314、電力起動リセット回路316、水晶発振器回路317、メモリ318、およびメモリ328を含む。SRS204はオプションとして、AC電源接続315、オプションのキーパッド311、またはオプションのディスプレイ312を含む可能性もある。ローカルなコマンド、プロンプトおよびデータなどの可聴発話の双方向的な通信のために、発話通信パスはVCC301、コーデック306、およびACI202を介する。非可聴発話通信パスは、遠隔コマンド、プロンプトおよびデータなどの、非可聴発話の双方向的な通信のために、VCC3

01、コーデック308、ECI207AまたはVCC301、ホスト・マイクロコントローラ310、APCC206、およびECI207Bを介する。ECI207は、電話ネットワーク、コンピュータ・ネットワーク、インターネット、高周波リンク、または赤外線リンクなどを介した有線リンクまたは無線リンクを提供することがある。

【0044】音声通信チップ301は音声制御式デバイス102に、本発明の標準の音声ユーザ・インタフェースを使用した発話を介した通信の能力を与える。マイクロフォン303は音声制御式デバイス102に、音声コマンドおよび装置の器具名などの可聴発話を聞く能力を与える。マイクロフォン303は用途に応じて、近距離音場マイクロフォンでも遠距離音場マイクロフォンでもよい。たとえば、近距離音場マイクロフォンは、ユーザの口が近い携帯セル電話内で好ましい可能性があり、一方、遠距離音場マイクロフォンはユーザの口が離れている車のセル電話内で好ましいことがある。スピーカ303を使用すると音声制御式デバイス102は、その名前またはコマンドの受信を確認するなどのために発話を使用して応答することができる。増幅器305は、マイクロフォン303によって受信された音声または発話信号を増幅する。さらに、増幅器305は人間のユーザ101が音声制御式デバイス102に正しくインタフェースできるように、スピーカ303を介したコーデック306からの音声信号表現を増幅できる。

【0045】マイクロフォン303とスピーカ304は各々、可聴発話と発話表現の間を変換するための変換器である。コーデック306は、ACI202からの発話表現を、VCC301のためのコード化された発話信号にコード化する。さらに、コーデック306は、VCC301からのコード化された発話信号を、ACI202を介した可聴通信のための発話表現に復号する。

【0046】別法として、非可聴発話信号は、音声制御式デバイス102によって双方向的に通信されることがある。この場合、VCC301はコード化された発話信号を復号するためにコーデック308に与える。コーデック308は、コード化された発話信号を復号し、接続105上の通信のためにそれをECI207Aに与える。発話信号は接続105上で受信され、ECI207Aに提供されることがある。ECI207Aは、コード化のために発話信号をコーデック308に結合する。コーデック308は発話信号をコード化し、そのコード化された発話信号がVCC301に結合される。

【0047】発話信号はまた、APCC206を介して電子的にも通信できる。送信のためのVCC301からの発話信号は、マイクロコントローラ310に渡される。マイクロコントローラ310はこれらをAPCC206に結合し、APCC206は発話信号をECI207Bへ送信する。音声制御式デバイス102によって受

信されるべき発話信号は ECI207B によって受信され、APCC206 に渡される。APCC206 は次いで、これらの受信された発話信号をマイクロコントローラ 310 に結合することがあり、マイクロコントローラ 310 は、これらの受信された発話信号を認識のために VCC301 上に渡す。

【0048】音声制御式デバイス 102 はホスト・マイクロコントローラ 310 からの信号によって、APCC206 を制御する。ホスト・マイクロコントローラ 310 は APCC206 に結合され、この制御を促進する。音声制御式デバイス 102 はオプションとして、さらなる入力手段として、マイクロコントローラ 310 に結合されたキーパッド 311 を有することがある。キーパッドは電力ボタン、会話ボタンへのプッシュ、または、セキュリティコード入力手段である可能性があり、さらに、オプションとして他の情報を入力するために使用される。音声制御式デバイス 102 はオプションとして、その状態またはユーザが関心のある他のアイテムを視覚的に表示するために、ホスト・マイクロコントローラ 310 に結合されたディスプレイ 312 を含む。しかし、音声制御式デバイスは一般に、オプションのキーパッド 311 またはオプションのディスプレイ 312 がなくても機能することができる。

【0049】音声制御式デバイス 102 は電源 314 を含む。電源 314 は、DC 供給源または AC 供給源、または両方から電力を生成してもよい。DC 供給源はバッテリー、太陽電池、または他の DC 源でもよい。AC 供給源の場合、オプションの AC 電力コード 315 が用意される。VCA102 は電力起動リセット回路 316 を含み、電源 314 が起動した時にそのシステムをリセットする。

【0050】水晶発振器回路 317 は VCC301 内の他の回路と共に、VCC301 に正確な振動入力を提供し、クロック信号を生成する。

【0051】メモリ 318 は VCC301 に結合され、書換え可能な非揮発性メモリおよび書換え可能な揮発性メモリ、および読み取り専用メモリである。これらは典型的にはフラッシュ RAM、静的 RAM、および ROM である。メモリ 318 は、プログラムを格納するため、およびあらかじめ記録されたフレーズおよび記録されたフレーズを格納するために使用される。さらに、メモリ 318 はプログラム動作のためにスクラッチ・メモリを設ける。業界内で標準の慣行であるように、使用されるメモリのタイプは、構成される特定の音声制御式デバイスによって変化することがある。本発明に関するプログラム格納は、ROM のように永久的である可能性、フラッシュのように非揮発性であるが可変的である可能性、RAM のように揮発性である可能性があり、この場合、プログラムは非揮発性メモリまたは遠隔の源からダウンロードされることがある。

【0052】メモリ 328 は揮発性メモリ、非揮発性メモリまたはその混合であることがある。揮発性メモリのみが使用されている場合、その内容は別の場所からダウンロードされて初期化できる。メモリ 328 の大きさと能力は、構成される音声制御式デバイスのタイプに依存する。別法として、場合によっては、メモリを磁気、光学、または他のタイプの記憶媒体で代用することができる。

【0053】音声制御式デバイス 102 内で、VCC301 はさらに、1 つの処理装置だけが音声制御式デバイス 102 内に含まれるようにホスト・マイクロコントローラ 310 の機能を含むことがある。同様に、統合レベルが増大し続けるにつれて当業界で慣習的であるように、APCC206、コーデック 306 および/または 308、ECI207A、ECI207B、メモリ 318、メモリ 328、増幅器 305、または他の要素がおそらく VCC301 に統合されるであろう。

【0054】次に図 4 を参照すると、音声通信チップ (VCC) 301 の構成図が示されている。音声通信チップ 301 は集積回路であり、処理装置 402、メモリ装置 403、バスとメモリ・コントローラ (BMC) 404、バス・アダプタ 405、周辺機器 406 を含む。音声通信チップ 301 はさらに、Information Storage Devices 社による「ISD-SR 300, Embedded Speech Recognition Processor」という名称のマイクロフィッシュの付録に説明されている。処理装置 402 はマイクロプロセッサとデジタル信号処理モジュール (DSPM) を含む。メモリ装置 403 は、DSPM ランダム・アクセス・メモリ (RAM) 407、システム RAM 408、読み取り専用メモリ (ROM) 409 を含む。周辺機器 406 は、I/O ポート 420、割込み制御装置 (ICU) 422、コード器/復号器 (コーデック) インタフェース 424、パルス幅変調器 (PWM) 426、MICROWIRE インタフェース 428、マスタ MICROWIRE コントローラ 430、リセットおよび構成コントローラ 432、クロック生成器 434 およびウォッチドッグ・タイマ 436 を含む。効果的に通信するために、音声通信チップ 301 は、コア・バス 415 および、図 4 に示されたような構成要素を内部接続する周辺バスを含む。

【0055】マイクロプロセッサ 416 は、RISKアーキテクチャを伴う汎用 16 ビット・マイクロプロセッサ・コアである。マイクロプロセッサ 416 は、整数の算術論理およびプログラム制御に責任を持つ。DSP モジュール (DSPM) 418 は DSP 算術を実行する。ROM 409 およびシステム RAM 408 は、プログラムとデータの格納のために使用される。DSPM RAM 407 は、DSPM 418 によって直接アクセスできる。DSPM 418 がアイドル状態の時、マイクロプロ

セッサ 416 は DSPM RAM 407 にアクセスできる。

【0056】バスとメモリ・コントローラ (BMC) 404 は、DRAM、拡張メモリ、オフチップ・ベース・メモリ、および I/O 拡張などのオフチップ措置へのアクセスを制御する。I/O ポート 420 は、音声通信チップ 301 に結合された装置へのインタフェースを提供する。I/O ポート 420 は、音声通信チップ 301 の 26 個の I/O ピンを表す。拡張オプションなしにプログラム・メモリに関して内部 ROM 409 を使用すると、16 個の I/O ピンが個別に入力用または出力用に構成でき、8 個の I/O ピンは出力専用にし、2 つの I/O ピンは入力専用にすることができる。ICU 422 は、5 個のマスク可能な割込み (4 つは内部的で 1 つは外部的) と、3 つの内部的なマスク不可能な割込み (NMI) を処理する能力を与える。コーデック・インタフェース 424 は、ACI 202 だけの場合は 1 つのコーデック装置 306 への直接インタフェースとなり、ACI 202 および ECI 207A の場合は 2 個のコーデック装置 306 および 308 への直接インタフェースとなる。パルス幅変調器 (PWM) 426 は、固定された周波数と可変デューティ・サイクルを有する方形波を生成する。MICROWIRE インタフェース 428 は、ホスト・マイクロコントローラ 310 とのシリアル通信を可能にする。マスタ MICROWIRE コントローラ 430 は、直列フラッシュ・メモリおよび他の周辺機器へのインタフェースを可能にする。リセットおよび構成ブロック 432 は、リセットの間、音声通信チップ 301 の環境のディフィニションを制御し、ソフトウェア制御構成を処理する。音声通信チップ 301 内の機能のいくつかは互いに排他的である。選択肢の間の選択は、リセットに際してまたはモジュール構成レジスタを介して行われる。クロック生成器 434 は、水晶発振器回路 317 にインタフェースし、リアルタイム・タイマを含む音声通信チップの種々のブロックのためにクロックを提供する。クロック生成器を使用し、音声通信チップ 301 をパワーダウン・モードに設定し、必要な時にそれを通常の動作モードに戻すことによって、電力消費を削減することもできる。音声通信チップ 301 がパワーダウン・モードである時、その機能のいくつかは使用不能になり、いくつかのレジスタの内容は変更されている。ウォッチドッグ・タイマ 436 は、ソフトウェアが処理ユニット 402 の制御を失った時はいつでも、および音声通信チップ 301 がパワーダウン・モードであった時には期間が終了した時にマスク不可能な割込みを生成する。

【0057】標準の音声ユーザ・インタフェース GUI を備えているコンピュータ動作システムと同様に、標準の音声ユーザ・インタフェース (VUI) も標準 VUI 動作システム・コードを備えていると考えることができ

る。広い範囲の多くの音声制御式デバイスで動作する標準 VUI を使用すると、ユーザは今まで対話できなかった装置を含めた任意の音声制御式デバイスとインタフェースできる。ユーザは一度標準の VUI に馴れると、標準の VUI で動作する任意の音声制御式デバイスに近づくすぐにその使用を開始できる。標準の VUI 動作システム・コードは、音声制御式デバイスを動作させる特定の標準化されたコマンドと手順を有する。これらの標準化されたコマンドと手順は、標準の VUI 動作システム・コードを実行するマシンに普遍的である。標準の VUI 動作システム・コードで動作する音声制御アプリケーション・ソフトウェアを書き、音声制御式デバイスを特定の用途にカスタマイズすることができる。音声制御アプリケーション・ソフトウェアは音声制御式デバイスが使用される用途に特有の音声コマンドを有する。特定の音声制御式デバイスはまた、標準 VUI の中心となる機能を拡張する追加の特別機能を備えても良い。

【0058】中心となる VUI の機能のうち標準の VUI 機能のいくつかは、音声制御式デバイスの存在を発見する方法、すべての音声制御式デバイスに関して共通のコア・コマンド・セット、音声制御式デバイスが応答するコマンドの種類を学ぶ方法 (コア・コマンドと器具特有のコマンドの両方)、マニュアルまたはディスプレイの使用なしにユーザを助けるための発音されたヘルプ・システム、ユーザが割当て可能な設定で音声制御式デバイスをユーザに個別化する方法、音声制御式デバイスの使用を許可されたユーザに制御し、ユーザが割当て可能な設定と情報を他のユーザから守るセキュリティ・メカニズム、ユーザが音声制御式デバイスと対話し、共通動作をするための標準の方法 (たとえばはいといいえの選択、選択肢のリストを一覧し、そこからアイテムを選ぶこと、優雅にエラーを処理する方法)。標準の VUI は API (アプリケーション・プログラミング・インタフェース) を含み、これを使用するとソフトウェア開発者は標準の VUI とインタフェースし動作して、音声制御コマンド・セットを拡張するカスタムの音声制御アプリケーションを書くことができる。

【0059】次に図 5 を参照すると、構成図は音声制御式デバイス 102 を制御し、標準の VUI および他の機能を提供するソフトウェア 500 を示す。ソフトウェア 500 はアプリケーション・コード 510、VUI ソフトウェア・モジュール 512、およびボキャブラリ 524 を含む。アプリケーション・コード 510 はさらに修正され、多数のアプリケーション・コードモジュールを表す複数のアプリケーションをサポートすることが可能であり、音声制御式デバイス 102 のさらなるカスタム化を提供する。ボキャブラリ 524 は検出されるべきフレーズを含む。ボキャブラリ内のフレーズはトピックと呼ばれるグループに分けられ、トピックは 1 つまたは複数あることがある。図 5 では、ボキャブラリ 524 はト



ピック 551 とトピック 552 の 2 つのトピックからなりたつ。

【0060】典型的に、アプリケーション・コード 510 は、アプリケーション・プログラミング・インタフェース (API) 507 を介して VUI ソフトウェア 512 にインタフェースする。VUI ソフトウェア 512 は、認識とプロンプト指示を含む音声インタフェースに関係するアプリケーション・コード 510 に、特別なサービスを提供する。VUI ソフトウェア 512 とアプリケーション・コード 510 の間の相互関係は、Microsoft 社の MS Windows と Microsoft Word の相互関係と類似している。Microsoft Windows は、画面上にアイテムを表示し、およびマウスの入力とキーボードの入力を受信することに関し、特別なサービスを Microsoft Word に提供する。

【0061】一般に、アプリケーション・コード 510 はホスト・メモリの中に格納され、ホスト・マイクロコントローラ 310 によって実行されることがある。しかし、ホスト・マイクロコントローラ 310 の機能は、ソフトウェア 500 に関連するコードを実行するために、1 つの装置またはプロセッサ、および 1 つのメモリまたは格納装置だけが必要とされるように、VCC 301 に埋め込まれることがある。

【0062】コア・コマンドとアプリケーションに特有のコマンドのためのフレーズを含め、認識可能なすべてのフレーズは、ボキャブラリ 524 に含まれる。VUI ソフトウェア・モジュール 512 は、たとえば認識の間に使用するために、ボキャブラリ・フレーズに直接アクセスできる。VUI ソフトウェア・モジュール 512 はまた、トークンを処理することもできる。トークンはトピック 551 ~ 552 内のフレーズに抽象的に結び付けられている。トークンは整数の数字である。たとえば「ダイヤルする」に関するフレーズは「5」のトークン値を有し、「電話を切ります」に関するフレーズは「6」のトークン値を有するといふようにである。認識できる各フレーズに 1 つのトークン値が割り当てられている。VUI ソフトウェア・モジュール 512 がボキャブラリ・ファイル 524 に関連したトークンを処理できるので、直接フレーズにアクセスする必要なくフレーズを参照することができる。これによって、VUI ソフトウェア・モジュール 502 を修正することなく、言語を変更する（英語からフランス語、など）ことができる。したがって、標準の VUI はボキャブラリ・ファイル 524 を修正するだけで、異なる方言や言語を使用して機能する。

【0063】ユーザは、音声制御式デバイス内で動作する標準 VUI の中心機能を使用すると、音声制御式デバイスを指名する、音声制御式デバイスの存在を識別する、ユーザが以前に格納した個別化された環境を起動

する、動作をキャンセルすることによって誤認識から回復する、ヘルプ機能を使用して音声制御式デバイスと共に使用できるコマンドと選択肢を識別する、標準的なコア・コマンドのセットおよび他の追加コマンドを使用する、コマンドが標準の構文に従っていることを確認することができる。（コマンドの構文は共通であるが、任意の音声制御式デバイス上のコマンドの特定のリストは、音声制御式デバイスの性格によって異なる）。標準の VUI はまた、API に関して次のユーザ対話のための標準機能を含む。GETYESNO—ユーザからのはい／いいえを受領する。GETRESPONSE—ユーザからの任意の入力を受領する。GETRESPONSEPLUS—強化されたエラー回復機能で、ユーザからの任意の入力を受領する。LISTANDSELECT—ユーザに選択のリストを提供し、ユーザに 1 つを選択させる。ACOUSTICADDDWARD—後から認識できるフレーズを追加する。

【0064】標準 VUI で正しく機能するために、音声制御式デバイス 102 の SRS 204 は、電力を供給された時に、会話とディジット（数）を連続的に認識する。しかし、一定の長さを越えた休止は、コマンドの終了をマークするまたは、不完全なコマンド・シーケンスが受信されたことの指示を与えるものとして SRS 204 によって認識されることがある。

#### 【0065】名前

本発明の標準 VUI の主要素は、各音声制御式デバイスが 1 つまたは複数の器具名を有し、器具名は各々がフレーズであるということである。最初の器具名は、工場ではメーカーによってプログラミングされた音声制御式デバイスのためのデフォルト名である。しかしユーザは一般に、音声制御式デバイスに彼らが選択したユーザ割当て器具名を割り当てることができる。音声制御式デバイスの命名は、人に命名するなどの別の種類の命名とは異なる。1 人の人は、彼と話したいすべての人に使用される単一の（ファースト）名を有する。対照的に音声制御式デバイスの命名では、音声制御式デバイスの各ユーザは通常、音声制御式デバイスに異なる、ユニークな名前を与える。したがって、音声制御式デバイスは、それが持つユーザの数と同じ数の名前を有することになるであろう。

【0066】ユーザが名前によって音声制御式デバイスを指名すると、2 つのことが起きる。第 1 に、音声制御式デバイスがその名前の 1 つを認識した時、音声制御式デバイスは装置が指名されたことを知らされ、コマンドを聞かなければならなくなる。第 2 に、各ユーザは通常 1 つの音声制御式デバイスについて異なる名前を使用するので、装置はユーザのアイデンティティ（発話者の識別）を知らされる。ユーザが音声制御式デバイスの機能に好みを有し、それを格納している場合、音声制御式デバイスはそのユーザの好みに合わせてそれ自体を個別に

設定することができる。

【0067】この命名のコンセプトを示すために、以下の例の、2人のユーザを有する音声制御式デバイスであるデスクトップ電話を考えてみる。ユーザ1は電話に「Aardvark」と命名し、ユーザ2は電話に「Platypus」と命名してある。電話が「Aardvark、母に電話します」と聞いた場合、電話はユーザ1によって指名され、およびユーザ1の電話帳を使用すべきであることを認識する。したがって、ユーザ1によってプログラミングされた「母」の番号をダイヤルする。同様に、電話が「Platypus、母に電話します」と聞いた場合、ユーザ2がそれを指名したことを知り、ユーザ2によってプログラミングされた「母」の番号をダイヤルする。

【0068】間違った認識を最小化するために、ユーザは音声制御式デバイスに、通常の会話では一般に話されない名前を割り当てることが好ましい。普通でない名前を選択することで、たがいに可聴範囲にある2つの音声制御式デバイスは同じ名前を有しない（おそらく異なるユーザによって割り当てられている）。音声制御式デバイスのメモリ制限があるため、いくつかの場合ではフレーズ名を話すための最大時間制限が要求されることがある。

【0069】次に、音声制御式デバイス102とともに標準VUIの詳細な動作のフロー・チャートが説明されている図6A～6Eを参照する。図6A～6Eのフロー・チャートでは、実線の箱はユーザによって通知されたフレーズ（引用符の中に置かれている）またはユーザの処置（引用符なし）を示す。点線の箱は、音声制御式デバイスによって通知されたフレーズ（引用符内）または取られた処置（引用符なし）を示す。点線の箱のすぐ下に直接実線の箱がある場合、現在の点線の箱内のアクションが正常に終了した場合に点線の右から出るパスがとられ、普通でないイベントが発生した場合には点線の箱の下にある実線の箱へのパスがとられる。一般に、点線の箱のすぐ下にある実線の箱は、普通でないイベントを示す。

#### 【0070】標準のVUIコマンド構文

次に図6Aを参照すると、すべての音声コマンドのための一般的な構文は、＜沈黙＞＜名前＞＜コマンド＞＜修飾子と変数＞である。＜沈黙＞は相対的な沈黙の期間で、そのあいだユーザは話さないが背景の騒音と背景の会話は依然として存在することがある。＜名前＞は音声制御式デバイス102に関連する器具名である。＜コマンド＞はユーザが実行したい動作である。＜修飾子と変数＞は、いくつかのコマンドによって必要とされる追加の情報からなりたつ。SRS204はユーザが音声制御式デバイスを制御できるように、構文の中にある要素を認識する。

【0071】ほとんどの音声制御式デバイスは連続的に

音声コマンド・シーケンスを聞く。音声制御式デバイスがその＜名前＞を聞く時、次の＜コマンド＞が指定されていることを知る。各ユーザが音声制御式デバイスに関して異なる＜名前＞を有するので、＜名前＞はまたユーザを一意的に識別し、音声制御式デバイスがそのユーザへの個別化を選択できる。コマンドは、すべての音声制御式デバイスに含まれるコアVUIコマンド、および所定のアプリケーションに特有のコマンドを含み、これらすべてはボキャブラリ524に格納されている。

【0072】＜名前＞の検出の前に＜沈黙＞を要求することは、通常の会話の間に＜名前＞の誤った検出を妨げる（つまり、ユーザが音声制御式デバイスに話しているのではなく、会話によって別のユーザに話している間）。すべての場合に、＜沈黙＞の長さはメーカによって構成され、0（＜沈黙＞を要求されない）から1秒またはそれ以上の範囲になることがある。典型的に、これは1秒の4分の1である。

【0073】Aardvarkと命名された電話などの音声制御式デバイスで使用される可能性のある音声コマンド・シーケンスの例は、「Aardvark、オフィスに電話します」「Aardvark、1-800-55-1212にダイヤルします」および「Aardvark、電話を切ります」を含む。（提供されたコマンドの例と説明では、簡潔にするために＜沈黙＞は示されることが多く、これが示されるまたは説明される時でも、メーカはゼロの長さの沈黙を使用するように選択する選択肢が常に存在する）。

【0074】コマンド構文が一般的な構文とは異なることが許される、2つの特別な場合がある。第1の特別な場合は、＜沈黙＞＜名前＞を連続的には聞いていない音声制御式デバイスの場合である。たとえば、いくつかのバッテリーで動作するアプリケーションの場合、電力消費に限度があるため、アイドル期間の間、音声制御式デバイス102内のVCC301が電力ダウンされることを要求する場合がある。別の例は、名前を間違って認識すると望ましくない結果になる場所にある音声制御式デバイスであり、たとえばプレゼンテーションの間の会議室内にあるデスクトップ電話である。第3の例は、たとえば多数の会話が聞かれる場所など、間違った認識のリスクが高い場所にある音声制御式デバイスである。

【0075】これらのタイプの状況については、代替のコマンド構文が所定の型のボタンまたはスイッチと共に使用される。第1の代替のコマンド構文は、＜スイッチの起動＞＜沈黙（オプション）＞＜名前＞＜コマンド＞＜修飾子と変数＞である。この構文では＜スイッチの起動＞は、ユーザがボタンを押すまたは所定の他の機械的な行動を実行して（たとえばフリップ型のセル電話を開ける）、認識機能を起動することを意味する。

【0076】第2の特別な場合は、ユーザが通常一連のコマンドを素早く連続して入力する場合である。これら



の場合については、ユーザはパスワード保護方法を使用し、または音声制御式デバイスの器具<名前>を含むコマンドを発行し、その後コマンドを続けて入力することによって、ユーザ自身を音声制御式デバイスに示すことができる。第2の代替のコマンド構文は（この例では、3つの連続的なコマンド）

<沈黙><名前><コマンド><必要に応じて修飾子と変数>

<沈黙><名前（オプション）><コマンド><必要に応じて修飾子と変数>

<沈黙><名前（オプション）><コマンド><必要に応じて修飾子と変数>

である。この構文を使用すると、ユーザは音声制御式デバイスの器具<名前>を常に繰り返す必要なく、一連のコマンドを発行できる。しかしユーザは、コマンドの開始時に<名前>を言うことが許されている。この構文では、<沈黙>は話された<名前>または<コマンド>を正しく認識するために要求されることに注意されたい。

【0077】第1または第2の代替構文のうちどちらかが使用されている場合、新しいユーザが音声制御式デバイスを使用し始めたかどうか、これらが正しく識別されているかどうかを確認することが望ましい。これは動作をしない時間の後、または音声制御式デバイスの電力をいれた後、または他の同様なプロトコルの後に、明示的に<名前>を要求することによって確認することができる。

#### 【0078】標準のコアVUIコマンド

標準VUIを使用して動作する音声制御式デバイス102のボキャブラリ524内に含まれる標準のコア・コマンドは、いくつかある。図6A～8は、以下のコマンドの構文を示す。

【0079】図6Aを参照すると、開始600の時点で音声制御式デバイスの器具名、<名前>は通常コマンドの前に話される。音声制御式デバイスがコマンドを聞いている時にいつでも音声制御式デバイスの任意の器具名を話すことができる。<名前>の後、所定の期間内にコマンドが続かない場合、音声制御式デバイスは元のアイドル状態の開始600に戻る。これはN秒の実線の箱の沈黙によって示される。この場合のNは通常、アプリケーションに依存し、音声制御式デバイスのメーカーによって割り当てられているプログラミング可能な値である。器具名を与えた後、601でユーザは音声制御式デバイス上で動作する標準のVUIのさらなるコマンドにアクセスすることを許可される。

【0080】ヘルプ・コマンドの構文は次の通りである。<名前>ヘルプ<コマンド（オプション）>、または、ヘルプ<コマンド（オプション）>。ヘルプ・コマンドは、任意の他のコマンドが与えられる時、または音声制御式デバイスが応答を待っている時いつでもをも含んだ任意の時に起動できる。音声制御式デバイスが有効

なコマンドを待っている間にヘルプ・コマンドが発行された場合、音声制御式デバイスが他のコマンドの前にく名前>を要求する場合は、ヘルプの前にく名前>がなければならない。音声制御式デバイスが任意の他のタイプの応答を待っている間にヘルプ・コマンドが要求された場合は、ヘルプ・コマンドの前にく名前>がある必要はない。<名前>がヘルプの前に要求されないすべての場合、ユーザが「<名前>ヘルプ」と言った場合、<名前>の使用はエラーを生じない。

10 【0081】ヘルプ機能は文脈依存であり、ヘルプが要求されるといつでも、音声制御式デバイスは音声制御式デバイスの現在の文脈が与えられれば、使用可能なオプションの説明で応答する。音声制御式デバイスがコマンドを聞いている時にヘルプが要求されている場合、音声制御式デバイスはその状態と、応答できるコマンドのリストで応答する（たとえば「メイン・メニューでは、「...」とすることができます」）。任意の特定のコマンドについてさらなる詳細は、「ヘルプ<コマンド>」構文（たとえば「ヘルプ・ダイアル」、「ヘルプ・コール」、「ヘルプ・ヘルプ」）で得ることができる。音声制御式デバイスが何らかの形の非コマンド応答（たとえば「名前を言う」）を待っている間に「ヘルプ」が要求された場合、音声制御式デバイスは音声制御式デバイスの現在の状態のステートメントで応答し、次に装置が何を待っているかの説明が続く（たとえば「ユーザの応答を待っています。あなたが作りたい電話帳エントリの人の名前を言ってください、またはNevermind（おかまいなく）と言ってキャンセルしてください」）。

30 【0082】キャンセル・コマンドの構文は次の通りである。<名前（オプション）>Nevermind、または、<名前（オプション）>キャンセルします。Nevermindコマンドまたはキャンセル・コマンドは、音声制御式デバイスがコマンドを実行し、ユーザからの応答を待っている時いつでも発行されるであろう。Nevermindまたはキャンセルは、音声制御式デバイスに現在のコマンドをキャンセルさせ、動作がキャンセルされたというステートメントで応答させる（たとえば「キャンセルされました」）。音声制御式デバイスがコマンドを待っている間にNevermindまたはキャンセルが発行された場合、これは無視される可能性もある。Nevermindまたはキャンセルを伴うく名前>の使用はオプションであり、これは<名前>が話されたかどうかとまったく同じように働く。

40 【0083】メイン・メニューへ戻るの構文は<名前>メイン・メニューである。コマンドのサブメニューを有する音声制御式デバイスについては、<名前>メイン・メニューはユーザをメイン・メニューに戻し、「メイン・メニューにいます」などの応答を起こさせる。このコマンドは、ユーザが任意のサブメニューから知られたボ

イントへ戻る簡単な方法を提供する。メイン・メニュー・コマンドは1つのメニューしか有しない音声制御式デバイス内では認識される必要はないが、サブメニューを伴う音声制御式デバイスについては必須のコマンドである。

#### 【0084】音声制御式デバイス名を変える

場合によっては、音声制御式デバイスのユーザ割当てされた名前を変更することが望ましい可能性もある。次に図6A～6Bを参照すると、名前変更コマンドの構文は次のとおりである。＜古い名前＞あなたの名前を変えます。このコマンドを使用すると、ユーザは音声制御式デバイスに命名したり命名を変更したりすることができる。音声制御式デバイスが新しい時、装置は少なくとも1つのデフォルトの、工場でプログラミングされた器具名を有する（たとえば電話）。ほとんどの音声制御式デバイスは、1つまたは複数のユーザが割当て可能な器具名をサポートする機能を有する。ユーザは、「＜工場でプログラミングされた名前＞あなたの名前を変えます」ということによって、器具名を命名することができる。

（たとえば「電話、あなたの名前を変えます」）。音声制御式デバイスは次いで、新しい名前が繰り返されることを求め、次いでその名前を変える。このプロセスは、各ユーザが割当て可能な名前についてもう一度繰り返すことができる。たとえば、ユーザが4人で、4つのユーザが割当て可能な器具名を割り当てることのできる電話を考えてみる。ユーザはコマンドで4つの名前の変更を実行することができる。「電話、あなたの名前を変えます」の次にユーザ1のための名前を（たとえば）Aardvarkに設定するダイアログが続く。「電話、あなたの名前を変えます」の次にユーザ2のための名前を（たとえば）Barracudaに設定するダイアログが続く。「電話、あなたの名前を変えます」の次にユーザ3のための名前を（たとえば）Coyoteに設定するダイアログが続く。「電話、あなたの名前を変えます」の次にユーザ4のための名前を（たとえば）Doggoneに設定するダイアログが続く。ユーザがコマンド（「電話、あなたの名前を変えます」）で続いて5番目のユーザが割当て可能な名前を変更しようと試みた場合、すべての使用可能なユーザが割当て可能な器具名が割り当てられているのでエラー・メッセージという結果になる。すべてのユーザ割当て名が定義されている場合でも、音声制御式デバイスは常に工場でプログラミングされた名前前で応答することに注意されたい。したがって、この第5の試みの例では、音声制御式デバイスは依然として工場でプログラミングされた「電話」という名前を認識するが、第5の新しいユーザが割当て可能な器具名を割り当てることはできないだけである。

【0085】既存のユーザが割当て可能な器具名はまた、「あなたの名前を変えます」コマンドで変更することもできる。上の例を続けると「Aardvark、あ

なたの名前を変えます」は第1のユーザに関する器具名を変更し（たとえば、これはPlatypusに変更できる）、他の3つのユーザ名を変更しないままに残す。同様に、「Platypus、あなたの名前を変えます」の次に名前を「電話」に変えるダイアログが続くと、第1のユーザ名を工場でプログラミングされたデフォルトにリセットする。

#### 【0086】音声制御式デバイスの識別

音声制御式デバイスが普及するにつれて、ユーザが新しい環境に入った時に音声制御式デバイスがそこに存在した場合、どの音声制御式デバイスが存在するかを容易に識別することができることが重要である。たとえば、ユーザがいくつかの装置を有するホテルの部屋に入る。これらを使用するために、ユーザはどの装置が音声制御式デバイスであるかを知る必要がある。さらに、ユーザは装置を正しく制御するために器具名を知る必要がある。可聴的に識別される他に、音声制御式デバイスは視覚的に識別され、また標準のVUIを使用する音声制御式デバイスを示すロゴを使用することによって識別できる。

【0087】音声制御式デバイスが起動して認識可能なコマンドを聞いている時、音響的な識別が機能する。ほとんどの場合、これは音声制御式デバイスが絶え間なく聞いて認識を試みていることを意味する。ほとんどのバッテリーで作動する音声制御式デバイスの場合は、この連続的な認識による電力の浪費は認められないため、これらの音声制御式デバイスではACで電力が供給されている。図6Aと図6Cを参照すると、音響的な識別はユーザが識別フレーズを通知し、音声制御式デバイスにコマンドすることによって達成できる。識別フレーズ「そこに何がありますか」、または何らかの他の適切な識別フレーズを使用して、音声制御式デバイスにそれ自体を認識させることが可能である。

【0088】標準のVUI識別フレーズの構文は、＜沈黙＞そこに何がありますか、である。この照会に回答して、この質問を聞いた任意の音声制御式デバイスが応答するはずである。典型的な音声制御式デバイスの応答は、最長で2秒までの相対的な沈黙のランダムな遅延であり、次にビーという音（標準の信号）、および「あなたは私を＜名前＞で呼ぶことができます」という応答が続く。ここで＜名前＞は、音声制御式デバイスを指名するために使用できる工場でプログラミングされた名前である。上に説明された電話の音声制御式デバイスの例では、応答は「＜ビー＞あなたは私を電話と呼ぶことができます」であろう。

【0089】図6Cを参照すると、最長で2秒までのランダムな遅延の間に、各応答する音声制御式デバイスは他の音声制御式デバイスの応答を聞く（特に、他の音声制御式デバイスのビーという音）。この沈黙期間の間に他の音声制御式デバイスが応答を始めた時（ビーという音によって明白である）、聞いている音声制御式デバ

スは応答している音声制御式デバイスが終了した後、その沈黙タイミングを再開始しなければならない。2つの音声制御式デバイスが同時に応答を開始した場合（ビーという音が重複した場合）、これらは両方とも新しくランダムに選択された沈黙遅延の間中断しなければならない。しかしこの時、ランダムな遅延は第1の遅延より長く、前の沈黙遅延の長さより最長で2倍程度である。どの場合でも、遅延は16秒を超えるべきではない。他の音声制御式デバイスが応答した場合、さらなる矛盾の解決のために追加の中断期間が与えられる。

【0090】図6Aを参照すると、ユーザが割当て可能な名前のリクエスト・コマンドの構文は、<名前>あなたの名前を教えてください、または<名前>あなたの名前（複数）を教えてください、である。セキュリティが許せば、任意のユーザがプログラミングした<名前>またはデフォルトの<名前>が使用できる。音声制御式デバイスに、装置が応答するすべてのユーザにプログラミングされた<名前>の一覧を求めるために、ユーザが割当て可能な名前のリクエスト・コマンドが使用される。セキュリティが許せば、音声制御式デバイスは、リスト様式で各ユーザがプログラミングした名前を通知する。各ユーザが割り当てた名前の中で、装置は一瞬一時停止する。この一時停止の間、ユーザはコマンドを音声制御式デバイスに通知する可能性があり、このコマンドは、音声制御式デバイスがそのユーザにプログラミングされた<名前>を与えられた時と同じように実行される。たとえば、上記の例の電話の音声制御式デバイスを考えてみる。一時停止の後にコマンド「電話、あなたの名前を教えてください」というコマンドは、「私はAardvark（一時停止）、Barracuda（一時停止）、Coyote（一時停止）、およびDoggone（一時停止）と命名されました」と言うことによって電話に回答させる。音声制御式デバイスが「Coyote」と言った後の一時停止の間、ユーザが「母に電話します」と言うと、この場合、電話はユーザCoyoteの母にコールする（母に関する電話番号がユーザCoyoteによってあらかじめ格納されていると仮定する）。

#### 【0091】セキュリティ上の考慮事項

ユーザが割当て可能な名前に関するコマンドは、音声制御式デバイスにおけるセキュリティの問題を提起する。いくつかの場合では、音声制御式デバイスへのアクセスを許可されたユーザに限定する必要がある。標準VUIによってサポートされている音声制御式デバイス内では、種々のセキュリティ保護の方法が使用できる。

【0092】もっとも簡単にセキュリティの程度が低い保護は、VUIの命名機能を介して与えられる。この場合、各ユーザは音声制御式デバイスについて一意的な名前を選択することが要求される。ユーザに割り当てられた器具名は音声制御式デバイス内で機密にされ、1人の

ユーザによってのみ変更され削除される。この方法では、器具名を使用して基本的なセキュリティを提供できる。しかし、この方法には多くの欠点がある。第1にユーザは典型的に、各コマンドを発行する前に名前を繰り返さねばならず、だれかが名前を簡単に立ち聞きできるようになり、セキュリティの損失という結果になる。第2に、ほとんどの音声制御式デバイスはその装置についてユーザ名を削除または変更する機能を含む。削除および変更を簡単に実行できるようにすることが望ましい。さらに、変更はその特定のユーザ以外のだれかによって実行される必要があることがある。たとえば、ユーザが音声制御式デバイスにもともと割り当てた名前を忘れる可能性もあり、またはそのユーザが装置の使用をやめ、彼の設定を削除できない可能性もある。器具名をセキュリティとして使用する場合は、簡単に名前を変えられるニーズと、セキュリティの質の間で固有の矛盾がある。

【0093】より高いレベルのセキュリティは、音声制御式デバイスへアクセスするときに、ユーザに秘密の数字のシーケンス、パスワード、またはフレーズを言うことを要求することによって達成されることがある。何らかの処置のない期間の後、または何らかの他の基準に基づいてユーザが音声制御式デバイスを使用する時に、ログインが要求されることがある。この方法の欠点は、話された数字のシーケンスまたはフレーズが立ち聞きされることがあることである。別のセキュリティの選択肢は、ユーザに数字のシーケンス、パスワード、またはフレーズを、オプションのキーパッド311などのキーパッド上で入力することを要求することである。これは追加のハードウェアを導入するが、これは別の人に秘密のコードを立ち聞きされるリスクを除去する。種々の他のセキュリティ・オプションもまた可能であり、その中には物理的なキーまたはセキュリティ・カードの使用も含まれる（たとえば磁気ストライプまたはスマートカード）。

【0094】追加のセキュリティは、音声制御式デバイスへのユーザのアクセスを自動的にキャンセルまたは終了することによって行われる。場合によっては、アクセスはコマンドの実行後ごとに、自動的にキャンセルされることがある。他の場合には、アクセスの自動的なキャンセルは、処置のない何らかの期間、電力ダウンまたはリセット、何らかの動作の完了（たとえば電話では、呼の終了時）の後、または「キャンセル・アクセス」コマンドの使用による特定の要求の時点で発生することがある。

#### 【0095】アプリケーション特有コマンド

標準のVUIは、各音声制御式デバイスにいくつかのアプリケーション特有コマンドを提供する。標準VUIによって提供されるアプリケーション特有コマンドは、電話と応答マシン・アプリケーションに関連付けられている。追加のアプリケーション特有コマンドはメーカーによ

ってそのボキャブラリ用にプログラミングされ、ボキャブラリの中に含まれることがある。

【0096】標準VUIのためにコマンドを開発する一般的なガイドラインは次の通りである。サブメニューは数を限定され、コマンドの論理的なグループの周囲に組織されるべきである。たとえば、電話TADは電話機能、電話帳管理のためのサブメニュー、およびTAD機能の他のサブメニューを含むメイン・メニューを有することがある。

【0097】任意のメニューまたはサブメニュー内のコマンドの数は一般に、10またはそれ以下に限定され、複雑さを最小限にするべきである。ヘルプ機能は明確に使用可能なコマンドを説明すべきである。

【0098】複雑なコマンドは、管理可能な小さな単位に分割すべきである。コマンドのフレーズは、認識の成功率を高くするように選択すべきである。標準のVUIコマンドは、認識の精度を高くするように選択すべきである。カスタム・ボキャブラリを作成する時、混乱をまねく恐れのあるフレーズの使用を避けるために注意をすべきである。

【0099】破壊的なイベント（削除など）に関しては、ユーザによる正しい入力の確認と動作の確認が要求されるべきである。

#### 【0100】電話のボキャブラリ

次に図6D～6E、図7、図8を参照すると、標準VUIのための電話ボキャブラリのためのフロー・チャートが示されている。電話ボキャブラリは特に、デスクトップ電話、セルラ電話、セルラ電話自動車キット、コードレス電話などの電話音声制御式デバイスのためのものである。本発明のSR S 204は電話ボキャブラリ内のコマンドを認識し、これらを電話音声制御式デバイスのための制御用に認識されたトークンに変換する機能を有する。電話ボキャブラリはすべての標準VUIコア・コマンドと、以下のアプリケーション特有コマンドを有する。

【0101】呼コマンドの構文は、<名前>電話します<ボイスタグ>、または<名前>電話します<ディジット>のどちらかである。呼コマンドを使用し、一連のディジットまたは電話帳ボイスタグのどちらかとして表現された特定の電話番号をダイヤルする。<ディジット>は数字のディジットの任意のリストであることがある。電話音声制御式デバイスは、ゼロに対して「oh（オー）」、ゼロ・ゼロに対し「hundred（ハンドレッド）」という別名が可能である。<ディジット>のシーケンスは埋め込まれた一時停止を含むことがある。しかし一時停止がプログラミング可能な長さを超えた場合、シーケンスは終了され、システムの設計者によって設定された長さを超えた一時停止を認識した後、コマンドが

実行される。呼コマンドに対する電話音声制御式デバイスの応答は、音声に表され、認識されたディジットまたは認識されたボイスタグを伴う、「<ディジット>に電話します」または「<ボイスタグ>に電話します」で、正しい認識を確認するべきである。誤認識の場合、「キャンセル」コマンドを使用して、呼動作をキャンセルすることがある。

【0102】ダイヤル・コマンドの構文は、<名前>ダイヤルします<ボイスタグ>、または<名前>ダイヤルします<ディジット>のどちらかである。ダイヤルコマンドは呼コマンドと同じである。

【0103】応答コマンドの構文は、<名前>応答します、である。このコマンドを使用して、着信呼に答える。応答プロンプトは「お話しください」である。

【0104】電話を切るコマンドの構文は、<名前>電話を切ります、である。このコマンドを使用して、起動中の呼の電話を切る。応答プロンプトは、ピッチの高いビーという音である。

【0105】再ダイヤル・コマンドの構文は、<名前>再ダイヤルします、である。このコマンドを使用して番号を再ダイヤルする。応答は「<ディジット>を再ダイヤルします」または「<ボイスタグ>を再ダイヤルします」であり、前の呼コマンドまたはダイヤル・コマンドが<ディジット>へのものであるかまたは<ボイスタグ>へのものであるのかによって異なる。それ以前に呼がなかった場合、応答は「再ダイヤルするものがありません」である。

【0106】格納コマンドの構文は、<名前>格納します、である。格納コマンドは電話帳サブメニュー内にあり、これを使用して新しいボイスタグを追加する。

【0107】削除コマンドの構文は、<名前>削除します、である。削除コマンドは電話帳サブメニュー内にあり、これを使用してボイスタグを削除する。

【0108】消音コマンドの構文は、<名前>消音します、である。このコマンドはマイクロフォンの音を消す。音声制御式デバイスによる応答は「消音されました」である。

【0109】オンライン・コマンドの構文は、<名前>オンラインにします、である。このコマンドはマイクロフォンの消音を外す。応答は「オンラインになりました」である。

【0110】ユーザからの応答を要求するために、プロンプトが音声制御式デバイスによって通知される。プロンプトは発話合成器、または、あらかじめ記録された発話の再生または他の手段によって通知される（すなわちプロンプト指示する）。電話ボキャブラリ内のプロンプトは次の、文脈依存ヘルプ・プロンプトを含む。

【表1】

「<ディジット><ボイスタグ>に電話します」	「電話したい名前を教えてください」	「オンライン」
「<ディジット><ボイスタグ>にダイヤルします」	「開始してください」	「1」
「お話しください」	「私の名前は<名前>になりました」	「2」
「さようなら（電話を切るコマンドに関して）」	「<ディジット><ボイスタグ>を再ダイヤルします」	「3」
「キャンセルされました」	「すみません、分かりません」	「4」
「削除したい名前を教えてください」	「名前をもう一度教えてください」	「5」
「<ボイスタグ>を削除してもよろしいですか」	「名前の変更がキャンセルされました」	「6」
「<ボイスタグ>が削除されました」	「名前が一致しません」	「7」
「新しい名前を教えてください」	「番号を繰り返してください」	「8」
「新しい名前を繰り返してください」	「<ボイスタグ>についての番号は<ディジット>です。よろしいですか」	「9」
「<ボイスタグ>について数字を教えてください」	「<ボイスタグ>についての番号が格納されました」	「0」
「その名前は電話帳にはありません」	「これを今格納しますか」	「100」
	「消音されました」	「再ダイヤルするものがありません」
		「星」
		「フラッシュ」
		「ボンド」

【0111】これらのプロンプトの他に、音声制御式デバイスはいくつかの異なるトーン音またはビーという音を生成できる。これらは中間ピッチのビーという音（たとえば500Hzの正弦波200ミリ秒間）、低いピッチのビーという音（たとえば間違った入力を示すブザーの音、または、低い周波数のビーという音250ミリ秒間）および高いピッチのビーという音（たとえば1200Hzの正弦波200ミリ秒間）を含む。他の音も可能であり、本発明の目的とする範囲の中である。

【0112】電話応答音声制御式デバイスのためのボキ

ャブラリ

前記の他に、標準VUIに関するアプリケーション特有コマンドを使用すると、ユーザは音声コマンドを使用している電話応答音声制御式デバイスへインタフェースできる。ユーザはキーパッドを使用しないでメッセージ機能を管理し、電話応答音声制御式デバイスからの遠隔アクセスを得ることができる。以下は、電話応答音声制御式デバイスについてボキャブラリ224に含まれるべき追加の音声コマンドを一覧したものである。

【表2】

<名前>新しく再生します	<名前>巻き戻しします<n>	<名前>停止します
<名前>すべて再生します	<名前>挨拶を記録します	<名前>挨拶を再生します
<名前>これを削除します	<名前>メッセージを記録します	<名前>部屋を監視します
<名前>すべてのメッセージを削除します	<名前>応答を起動します	<名前>パスワード
<名前>正方向へ進みます<n>	<名前>応答を終了します	<パスワード・フレーズ>

### 【0113】自動車制御ボキャブラリ

標準のVUIに関する追加の特定のコマンドを使用すると、ユーザは音声制御を使用した自動車のアクセサリにインタフェースできる。自動車の音声制御に関する2つの主な領域は、車内アクセサリの制御と、娯楽システムの制御を含む。自動車のアクセサリは環境の制御、ウィンドウ、ドア・ロック、車内の照明を含む。ステアリング、ブレーキ、加速、および車外の照明などの自動車内の「ミッション・クリティカル」な要素は、誤認識が発生した時に安全面での心配が起きることがあるため、音声によって制御されないことが好ましい。娯楽の制御は第1にCDプレイヤー/チェンジャおよびラジオのために使用される。

【0114】音声制御式デバイスのための自動車制御ボキャブラリ224は、空気調整、ファンの速度、温度、ドライバのウィンドウ、乗客のウィンドウ、左の後部ウィンドウ、右の後部ウィンドウ、ウィンドウ、ドア・ロック、ワイパ、低い、中間、高い、増加、減少、設定、再設定、キャンセル、消去、再呼び出し、オン、オフ、より涼しく、およびより暖かくを含む。

【0115】APIに関する標準のユーザ・インタフェース機能

本発明の標準VUIはユーザの対話のための標準機能を含み、これはアプリケーション・プログラミング・インタフェース(API)によってアクセスされる。これらのAPIに関する標準機能はGETYESNO、GETRESPONSE、GETRESPONSEPLUS、およびLISTANDSELECTを含み、これらはカスタム・ソフトウェア開発者によって使用されて、本発明の標準VUIの上で動作するアプリケーションを開発する。図9A～9B、図10A～10C、図11、および図12は、標準VUI内のこれらの標準ユーザ・インタフェース機能の機能を示すフロー・チャートである。簡単に言えば、GETYESNO機能はユーザから肯定的な応答(はい)、または否定的な応答(いいえ)をプロンプト指示し、受け入れるためのものである。GETRESPONSE機能は、期待される応答のリストに対応するユーザからの入力をプロンプト指示し、受け入れるためのものである。GETRESPONSEPLUS

機能はGETRESPONSE機能と同様にユーザからの入力をプロンプト指示し、受け入れるためのものであるが、強化されたエラー回復機能を有する。LISTANDSELECT機能はユーザに選択肢の一覧を提供し、ユーザが選択肢の1つを選択できるようにする。GETYESNO、GETRESPONSE、GETRESPONSEPLUS、およびLISTANDSELECTの動作は、1997年のInternational Journal of Speech Technologyにある、Bruce E. Balentineらによる、「Debouncing the Speech Button: A sliding Capture Window Device for Synchronizing Turn-Taking」から改良した。図9Aは、はい/いいえメニューの使用を示し、図9Bは、拒否または間違った認識を解決する方法を示す。図10Aは、GETRESPONSE機能およびGETRESPONSEPLUS機能に関して、始動ウィンドウまたは開始ウィンドウを示す。図10Bは、GETRESPONSE機能およびGETRESPONSEPLUS機能に関して、会話開始ウィンドウまたは開くウィンドウ機能を示す。図10Cは、GETRESPONSE機能およびGETRESPONSEPLUS機能に関して、認識終了ウィンドウまたは閉じるウィンドウ機能を示す。図11は、GETRESPONSEPLUS機能に関して二重の取込みウィンドウを示す。図12は、LISTANDSELECT機能に関してメニュー・リスト機能を示す。

【0116】図9A～9Bを参照すると、GETYESNOユーザ・インタフェース機能を使用すると、ユーザに質問をし、「はい」または「いいえ」(または他の言語内の同等のフレーズ)などの肯定的な応答または否定的な応答を受け取る。GETYESNOに関連するパラメータは、QUESTION期間とTIMEOUT期間である。質問パラメータはユーザへの音声プロンプトであり、「はい」または「いいえ」などで肯定的または否定的に応答できる質問をする。タイムアウト・パラメータは応答が検出されなかったとフラグを付ける前に応答を待つ秒数である。音声制御式デバイスは応答または結

果に応じて、バイト値を戻す。「いいえ」応答が検出された場合は、0が戻される。「はい」応答が検出された場合は、1が戻される。許可された時間内に応答が検出なかった場合は、TimeOutエラーを示す17が戻される。応答が検出されても認識不可能で、ボキャブラリ外の単語エラーを示す場合に、18が戻される。

【0117】図10A~10Cを参照すると、GETRESPONSEユーザ・インタフェース機能は応答を促し、応答を待つPromptをユーザに再生する。GETRESPONSEは、TopicListとして知られるリスト内のトピックに一致する話された応答を探す。GETRESPONSEは、認識されたトークンのアレイを戻すか、エラー標識を戻す。GETRESPONSEに関連するパラメータは、Prompt、TimeOut、STS\_\_SoundおよびTopicListである。Promptパラメータは、ユーザに送信されるべき最初のプロンプトである。TimeOutパラメータは、応答が検出されなかったとフラグを付ける前に応答を待つミリ秒の数である。STS\_\_Soundパラメータ（早すぎた発話の音）は、ユーザがPromptの再生が終わる前に話した場合、再生される音またはプロンプトである。典型的には、STS\_\_Soundは話されたフレーズではなく短いトーン音またはビーという音である。パラメータTopicListは、SRS204が話された応答を識別するために使用すべきトピックのリストに関するボキャブラリ・サブセットである。音声制御式デバイスは整数アレイへのポインタを戻す。TopicListに関連する応答の認識が成功した場合、アレイ内の最初の要素は戻されたトークンの数であり、アレイ内の続く要素は各識別された発話要素に関するトークン（1または複数の単語）である。要素1はn、戻されるトークンの数である。要素2からn+1は、認識された各発話要素に関するToken値である。たとえば、「電話、オフィスにダイヤルします」というフレーズを考えてみる。発話要素「電話」のためのトークン値が7、発話要素「ダイヤル」が12、発話要素「オフィス」が103であった場合、次いで、これらがすべてうまく認識された場合、戻される完全なアレイは値3、7、12、103を伴う4要素長である。応答の認識が成功しない場合、アレイは2要素長である。最初の要素はゼロに設定され、第2の要素は発生したエラーのタイプを示す。この場合要素1は0に設定され、エラーが検出されたことを示す。要素2は17に設定され、応答が許可された時間内で検出されなかったことを示すか（TimeOutエラー）、または18に設定され、応答が検出されたが認識不可能であったことを示す（ボキャブラリ外の単語のエラー）。タイムアウト・エラーに関して戻されるアレイは、値0と17を伴う2要素長であり、ボキャブラリ外の単語エラーに関して戻されるアレイは、値0と18を伴う2要素長である。

【0118】図11を参照すると、GETRESPONSEPLUSユーザ・インタフェース機能はユーザにPromptを再生し、応答を促し、応答を待つ。GETRESPONSEPLUSは、ユーザに対してPromptを再生し、次いで話された応答を待つという点でGETRESPONSEと同様である。しかし、GETRESPONSEPLUSは、プロンプトを再生し、ユーザが話さなかったり、または背景に過剰な雑音を有するエラー状況から回復するプロンプトを再生する機能を有する。GETRESPONSEPLUSは、TopicList内のトピックに一致する話された応答を聞く。GETRESPONSEPLUSは、認識されたトークンのアレイか、エラー標識のどちらかを戻す。GETRESPONSEPLUSのパラメータはInitial\_\_Prompt、Timeout、STS\_\_Sound、TopicList、MaxTries、Intervene\_\_Prompt、Repeat\_\_Prompt、およびHelp\_\_Promptである。Initial\_\_Promptパラメータは、ユーザに再生され応答を促すべき最初のプロンプトである。Timeoutパラメータは、応答が検出されなかったとフラグを付ける前に応答を待つのためのミリ秒の数である。STS\_\_Soundプロンプトは、Promptの再生が終了する前にユーザが話した場合に再生されるべき音またはプロンプトである。典型的には、STS\_\_Soundプロンプトは話されたフレーズではなく、短いトーン音またはビーという音である。TopicListパラメータは、SRS204が使用して話された応答を識別すべきトピックのリストに関するボキャブラリ・サブセットである。MaxTriesパラメータは、GETRESPONSEPLUSがよく認識しようとして、ユーザに再びプロンプト指示する最大の回数である。MaxTriesの後認識できない場合、GETRESPONSEPLUSが戻り、エラーを示す。Intervene\_\_Promptパラメータは、ユーザに繰り返しを求めるために再生されるプロンプトである（たとえば「雑音が多すぎます。言ったことを繰り返してください」）。このプロンプトは、前の認識の試みの間に雑音が多すぎた場合に、再生される。Repeat\_\_Promptパラメータは、言ったばかりのことの繰返しをユーザに求めるために再生されるプロンプトである（「言ったことを繰り返してください」など）。このプロンプトは、発話が早すぎるエラーが発生した時に使用される。Help\_\_Promptパラメータは、ユーザがさらに命令を必要とするように見える時に再生されるプロンプトで、ユーザが何も言わない時も含まれる。音声制御式デバイスは、ユーザ・インタフェース機能の終了時点で、整数アレイへのポインタを戻す。TopicListに関連する応答の認識が成功した場合、アレイ内の第1の要素は戻されたトークンの数であり、アレイ内の続く要素は各



識別された発話要素に関するトークンである（1つまたは複数の単語）。要素1はnで、戻されたトークンの数である。要素2からn+1は、認識された各発話要素に関するトークン値である。たとえば「電話、オフィスにダイヤルします」というフレーズを考えてみる。発話要素「電話」のトークン値が7であり、発話要素「ダイヤル」のトークン値が12、発話要素「オフィス」のトークン値が103であり、これらがすべてうまく認識された場合、戻される完全なアレイは4要素長で値は3、7、12、103である。認識が成功しなかった場合、アレイは4要素長である。第1の要素はゼロである。第2の要素は発生したもっとも最近のエラーのタイプを示す。第3の要素から第5の要素は、GETRESPONSEPLUSが呼ばれた時からGETRESPONSEPLUSが戻った時までの間に発生した、各タイプのエラーの回数を示す。この場合、要素1は値0を有し、エラーが検出されたことを示す。要素2は値17を有し、応答が許可された時間内に検出されなかったことを示すか（TimeOutエラー）、値18を有し、応答が検出されたが認識不可能であったことを示し（ボキャブラリ外の単語エラー）、または値19を有し、早すぎた発話エラーが検出されたことを示す。要素3は値xを有し、TimeOutエラーが検出された回数を示す。要素4は値yを有し、ボキャブラリ外の単語エラーが検出された回数を示す。要素5は値zを有し、早すぎた発話エラーが検出された数を示す。

【0119】図12を参照すると、LISTANDSELECTユーザ・インタフェース機能は第1にPromptを再生する。次いで、ListOfMenuPromptsアレイ内で各プロンプトを再生し、各プロンプトの後にPauseTimeの間、一時停止する。これらの一時停止の間、認識器はTopicList内のトピックに一致する話された応答を聞く。LISTANDSELECTは認識されたトークンのアレイか、エラー標識のどちらかに戻る。LISTANDSELECTに関するパラメータは、Initial\_Prompt、Timeout、STS\_Sound、TopicList、ListOfMenuPrompts、PauseTime、およびHelp\_Promptを含む。Initial\_Promptパラメータは、ユーザに再生されるべき最初のプロンプトである。Timeoutパラメータは、ListOfMenuPrompt内のすべてのプロンプトを再生した後、または応答が検出されなかったとフラグを付ける前に、応答を待つためのミリ秒数である。STS\_Soundパラメータは、プロンプトの再生を終わる前にユーザが話した場合に再生されるべき音またはプロンプトである。典型的には、STS\_Soundは話されたフレーズではなく短いトーン音またはビーという音である。TopicListパラメータはSRS204が使用して話された応答を識別す

べき、トピックのリストに関するボキャブラリ・サブセットである。ListOfMenuPromptsパラメータは、一度に1つ再生されるプロンプトのアレイである。アレイ内の第1の要素は、ListOfMenuPrompts内のプロンプトの数のカウントである。PauseTimeパラメータは、ListOfMenuPrompts内の各プロンプトを再生した後に一時停止する時間である。PauseTimeパラメータは、ミリ秒の値を有する。Help\_Promptパラメータは、ユーザが何も言わなかった時を含め、ユーザがさらなる命令を必要とするように見える時に再生されるプロンプトである。音声制御式デバイスはユーザ・インタフェース機能の完了時に、整数アレイへのポインタを戻す。認識が成功した場合、アレイ内の第1の要素は戻されたトークンの数であり、アレイ内の続く要素は各識別された発話要素に関するトークンである（1つまたは複数の単語）。要素1は値nを有し、戻されたトークンの数を示す。要素2からn+1は値xを有し、認識された各発話要素に関するトークン値を示す。認識が成功しなかった場合、アレイは2要素長である。第1の要素はゼロである。第2の要素は発生したエラーのタイプを示す。この場合、要素1は値0を有し、エラーが検出されたことを示す。要素2は値17を有し、応答が許可された時間内に検出されなかったことを示すか（TimeOutエラー）または、値18を有し、応答が検出されたが認識不可能であったことを示す（ボキャブラリ外の単語エラー）。

【0120】ACOUSTICADDDWORD機能はアプリケーション・ソフトウェアによって使用され、ユーザがフレーズ、またはボイスタグと呼ばれるものを音声制御式デバイスに追加することを可能にする。これらのフレーズはGETRESPONSE機能およびGETRESPONSEPLUS機能を使用して後から認識できる。ACOUSTICADDDWORD機能を使用して、たとえば、電話内で名前によるダイヤル入力を作成することができる。人の名前（「ジョン・スミス」）またはアイデンティティ（「母」）または他の区別するフレーズ（「私のオフィスの番号」）をACOUSTICADDDWORDで格納することにより、人は「ジョン・スミスに電話します」「母に電話します」、または「私のオフィス番号に電話します」と言うことによってその番号に後から電話することができる。

【0121】ACOUSTICADDDWORDは、ボイスタグを指定されたTopicListに格納する。動作において、ACOUSTICADDDWORDはプロンプトを再生し、ボイスタグを受信し記録し、ボイスタグを確認し、次いでボイスタグを格納する。AcousticAddWordは、ボイスタグを複数回再チェックすることによってエラーから回復する能力を有する。AcousticAddWordは、重複の場合チェック



し、エラーをユーザに戻す。ACOUSTICADDWORDに関するパラメータは、Initial\_Prompt、Timeout、STS\_Sound、TopicList、MaxTries、Repeat\_Prompt、Intervene\_Prompt、Error\_Prompt、Ok\_Prompt、およびHelp\_Promptを含む。Initial\_Promptパラメータはユーザに再生されるべき最初のプロンプトであり、音声制御電話の電話帳に名前を格納する例では「新しい名前を教えてください」などである。Timeoutパラメータは、失敗が検出されたというフラグを応答に付ける前に待つミリ秒数である。STS\_Sound (Spoke-Too-Soon Sound) パラメータは、Promptが再生を終了する前にユーザが話した場合に再生されるべき音またはプロンプトである。典型的に、STS\_Soundは話されたフレーズではなく短いトーン音またはビーという音である。パラメータTopicListはSRS204が新しいボイスタグを格納すべきボキャブラリ・サブセットである。MaxTriesパラメータはよく認識しようとして、AcousticAddWordがユーザに再びプロンプト指示する最大数である。MaxTriesの後、認識できない場合、AcousticAddWordはエラー指示を戻す。Repeat\_Promptパラメータは、言ったばかりのことの繰返しをユーザに求めるために再生されるプロンプトである（たとえば「言ったことを繰り返してください」）。このプロンプトは、発話が早すぎるエラーが発生した時に使用される。Intervene\_Promptパラメータは、ユーザに繰返しを求めるために再生されるプロンプトである（たとえば「雑音が多すぎます。言ったことを繰り返してください」）。このプロンプトは、前の認識の試みの間に雑音が多すぎる時に再生される。Error\_Promptパラメータは、繰り返された名前が最初の名前に一致しない時か、または名前が二重であった場合に再生されるプロンプトである（たとえば「もう一度試みてください」）。OK\_Promptパラメータは、新しい名前がうまく記録され格納された時に再生されるプロンプトである（たとえば「<名前>がアドレス・ブックに格納されました」）。Help\_Promptパラメータは、ユーザが何も言わなかった時を含め、ユーザがさらなる命令を必要とするように見える時に再生されるプロンプトである。音声制御式デバイスは、ユーザ・インタフェース機能の完了時に、整数アレイへのポインタを戻す。AcousticAddWordに関連する応答の認識が成功した場合、アレイは7要素長である。要素1は値1であり、成功した認識を示す。要素2はSRS204によって割り当てられたトークン数を示す値であり、格納されたボイスタグに対応する。要素3は、ボイスタグの記録されたコピーへのポインタであ

る。要素4は、発生したタイムアウト・エラーの回数を示す値である。要素5は、名前に一致しなかった失敗があった回数を示す値である。要素6は、早すぎる発話が発生した回数を示す値である。要素7は、ヘルプ・プロンプトが再生された回数を示す値である。認識が成功しなかった場合、アレイは6要素長である。第1の要素はゼロである。第2の要素は発生したエラーのもっとも最近のタイプを示す。第3の要素から第5の要素は、AcousticAddWordが呼ばれた時からAcousticAddWordが戻った時までの間に発生した各タイプのエラーの回数を示す。第6の要素は、ヘルプ・プロンプトが再生された数である。この場合、要素1はエラーが検出されたことを示す値である。要素2は、値17を有し、応答が許可された時間内に検出されなかったことを示し（Timeoutエラー）、値18は応答が検出されたが認識不可能であることを示し（Noiseエラー）、値19は発話が早すぎるエラーが検出されたことを示し、値20はRecognitionの失敗を示し（繰返しに一致がない）、または値21はVoicetagリストがすでに一杯であることを示す。要素3は値xで、Timeoutエラーが検出された回数を示す。要素4は値yで、認識エラーが検出された回数を示す。要素5は値zで、発話が早すぎるエラーが検出された回数を示す。要素6は、ヘルプ・プロンプトが再生された回数を示す値である。

【0122】音声制御式デバイスに関するエチケット標準VUIは、音声制御式デバイスに関するエチケットを含む。一般に、音声制御式デバイス（マシンとも呼ばれる）は、行儀のよいお客のようにふるまうべきである。

【0123】しかし、音声制御式デバイスと一緒の生活に含まれる人間的な要素と人間的な問題はあまり探索されていない。音声制御式デバイスの設計の際は、次の提案が考慮されなければならない。

【0124】人間に要求するマシン

マシンは人間に何かするように求めることができる。任意の要求はていねいであるべきである。たとえば音声起動セルラ電話は、そのバッテリーが低くなった時に充電することを求めることがある。人間はつねにマシンの要求を拒否する選択肢を有し、マシンは人間の生命または貴重なデータを脅かすような状況であると考えない限り、それをていねいに受け入れるべきであり、脅かすような状況である場合はさらに緊急な抗議をすることがある。

【0125】自分自身のために電話を使用するマシン  
音声制御式デバイスが電話に応答した場合、または人間のユーザに電話した場合、その呼が人間と考えられるリスクがあった場合には、マシン自体でマシンであると自分を明確にすべきである。

【0126】ユーザの発話の記録

どのマシンも、そこにいる人間が会話が記録または転写

されていることを認識しているものでなければ、人間のユーザの会話を記録または転写すべきではない。

#### 【0127】ボリューム・レベル

マシンは、明確に人間の声の方が大きいのでない限り、周囲の騒音レベルに応答してボリューム・レベルを変調すべきである。マシンは人間が静かにして欲しいと思う時にはそれに敏感であるべきである（たとえば人間が眠っている時）。マシンは不必要にしゃべるべきではなく、これらを黙らせる手段としてユーザが介入できるようにすべきである。

#### 【0128】マシンからマシンへの通信

図13は、通信する1対の音声制御式デバイス102Mと102Nの構成図である（各々はマシンとも呼ばれる）。その一方または両方は通信環境1300内で本発明の標準の音声ユーザ・インタフェース500を使用することができる。音声制御式デバイスは互いに話して、他にどの音声制御式デバイスが存在するか、これらがどの種類の情報を理解するかを見つけ出し、情報を交換することができる。たとえば、音声制御TVは音声制御VCRに、動作するために必要な設定に関して尋ねることがある。音声制御式デバイス間のマシンからマシンへの通信は、可聴フォーマットおよび非可聴フォーマットの両方で発生する。本質的に、発話を使用したマシンからマシンへの通信は任意の発話互換媒体上で発生する可能性があり、空気を介した音波、従来の電話リンク、インターネットの音声リンク、無線の音声チャネルなどを含む。マシンからマシンへの通信は、マシンの一部、またはマシンのすべてが本発明のVUIを含む場合にも起き、またどのマシンも本発明のVUIを含まない場合にも起きる。

【0129】標準のVUIを使用すると、音声制御式デバイスは他の音声制御式デバイスを多くの方法で通信環境内に置くことができる。これらは人間の他のマシンとの対話の立ち聞き、マシンの別のマシンとの対話の立ち聞き、識別フレーズ「＜沈黙＞そこに何がありますか」を使用することによって自分自身を識別するように近くのマシンに明示的に要求すること、「＜沈黙＞時計、そこにありますか」という名前カテゴリーによってそれらを示すことによって特定の種類のマシン（たとえばすべての時計）を明示的に探すこと、または「＜沈黙＞Socrates、そこにありますか」という名前によって指名して特定のマシン（たとえば、Socratesと名付けられたクロック）を明示的に探すことを含む。

【0130】最初の2つの場合では、他の会話を聞くプロセスは、別のマシンの名前を明らかにする。別の3つの場合は、「あなたはそこにありますか」コマンドに回答した、呼べば聞こえる場所にいるマシンは、自分の名前前で応答する。最後の2つの場合は、「そこに何がありますか」コマンドは、マシンの所定の種類と、特定の名前のマシンに限定され、これによってそのコマンドに

答するマシンの数を限定する。目的の音声制御式デバイスの名前が分かると、最初の音声制御式デバイスは別のコマンドを別のマシンに発行することができる（たとえば「Socrates、今何時ですか」）。

【0131】場合によっては、音声制御式デバイスは別の音声制御式デバイスに話しかける必要があることがあるが、その装置のうちどちらか1つかまたは両方とも、上記のプロトコルを厳守していないことがある。これらの場合、マシンを明示的にプログラミングして正しいコマンドを発行し、適切な応答を認識することができる。この対話の簡単な例は、音声制御能力を有する音声制御式デバイス、および、発話された時間報告や、所望のデータ（時間）をただ捕捉するなどの音声に基づいたサービスをダイアルする電話音声インタフェースである。

【0132】したがって、音声制御デバイスの活動化を向上させる方法及び装置に関する、本発明の好ましい実施形態が説明される。本発明の好ましい実施形態は発話者から独立した音声認識システムを使用しているが、本発明はまた、発話者に依存した音声認識システムとも互換である。本発明が特定の実施形態内で説明されたが、本発明はこのような実施形態によって限定されると解釈されるべきではなく、首記の請求の範囲に従って解釈されるべきである。

#### 【図面の簡単な説明】

【図1A】本発明の音声制御式デバイスを含む環境を示す図である。

【図1B】図1Aに示された環境内における、音声制御式デバイスでの遠隔通信を示す図である。

【図2】例としての音声制御式デバイスの図である。

【図3】本発明の音声制御式デバイスの詳細な構成図である。

【図4】音声通信チップの詳細な構成図である。

【図5】本発明の標準の音声ユーザ・インタフェースの構成図である。

【図6A】～

【図6C】本発明の標準の音声ユーザ・インタフェースに関するコア・コマンド構造のフロー・チャートである。

【図6D】～

【図6E】本発明の標準の音声ユーザ・インタフェースに関する電話コマンド構造のフロー・チャートである。

【図7】本発明の標準の音声ユーザ・インタフェースに関する「名前を格納する」電話コマンド構造のフロー・チャートである。

【図8】本発明の標準の音声ユーザ・インタフェースに関する「名前を削除する」電話コマンド構造のフロー・チャートである。

【図9A】～

【図9B】本発明の標準の音声ユーザ・インタフェースに関する「GETYESNO」機能のフロー・チャート

である。

【図10A】～

【図10C】本発明の標準の音声ユーザ・インタフェースに関する「GETRESPONSE」機能のフロー・チャートである。

【図11】本発明の標準の音声ユーザ・インタフェースに関する「GETRESPONSEPLUS」機能のフロー・チャートである。

【図12】本発明の標準の音声ユーザ・インタフェースに関する「LISTANDSELECT」機能のフロー・チャートである。

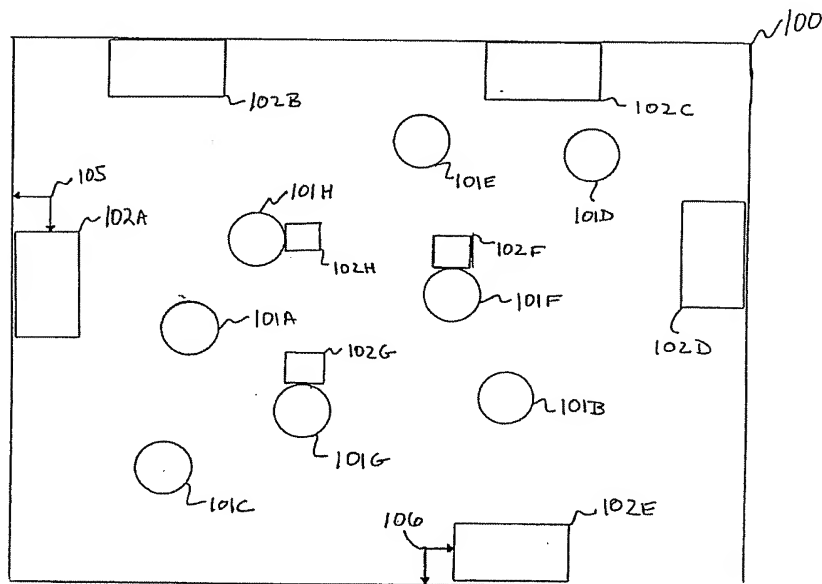
【図13】本発明の標準の音声ユーザ・インタフェースを使用して通信する一対の音声制御式デバイスの構成図

である。図の中の同様な参照番号と指示は、同様な機能を提供する同様な要素を示す。

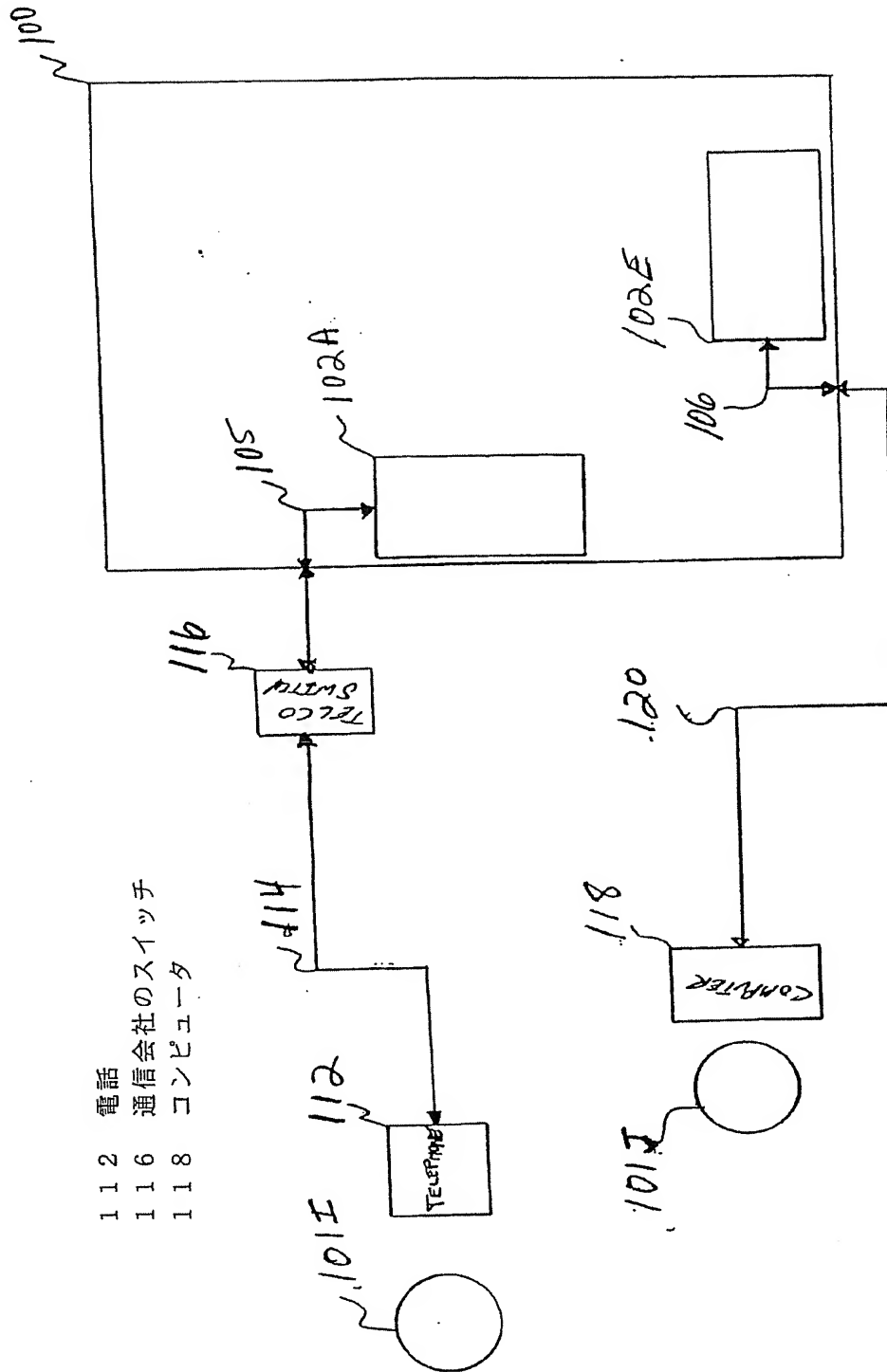
【コードの説明】

- 100 環境
- 101A～101H 人間のユーザ
- 102A～102H 音声制御式デバイス
- 105 接続
- 112 電話
- 114 伝送媒体
- 116 電話会社のスイッチ
- 118 遠隔コンピュータ
- 120 ネットワーク接続
- 102I～102M 音声制御式デバイス

【図1A】

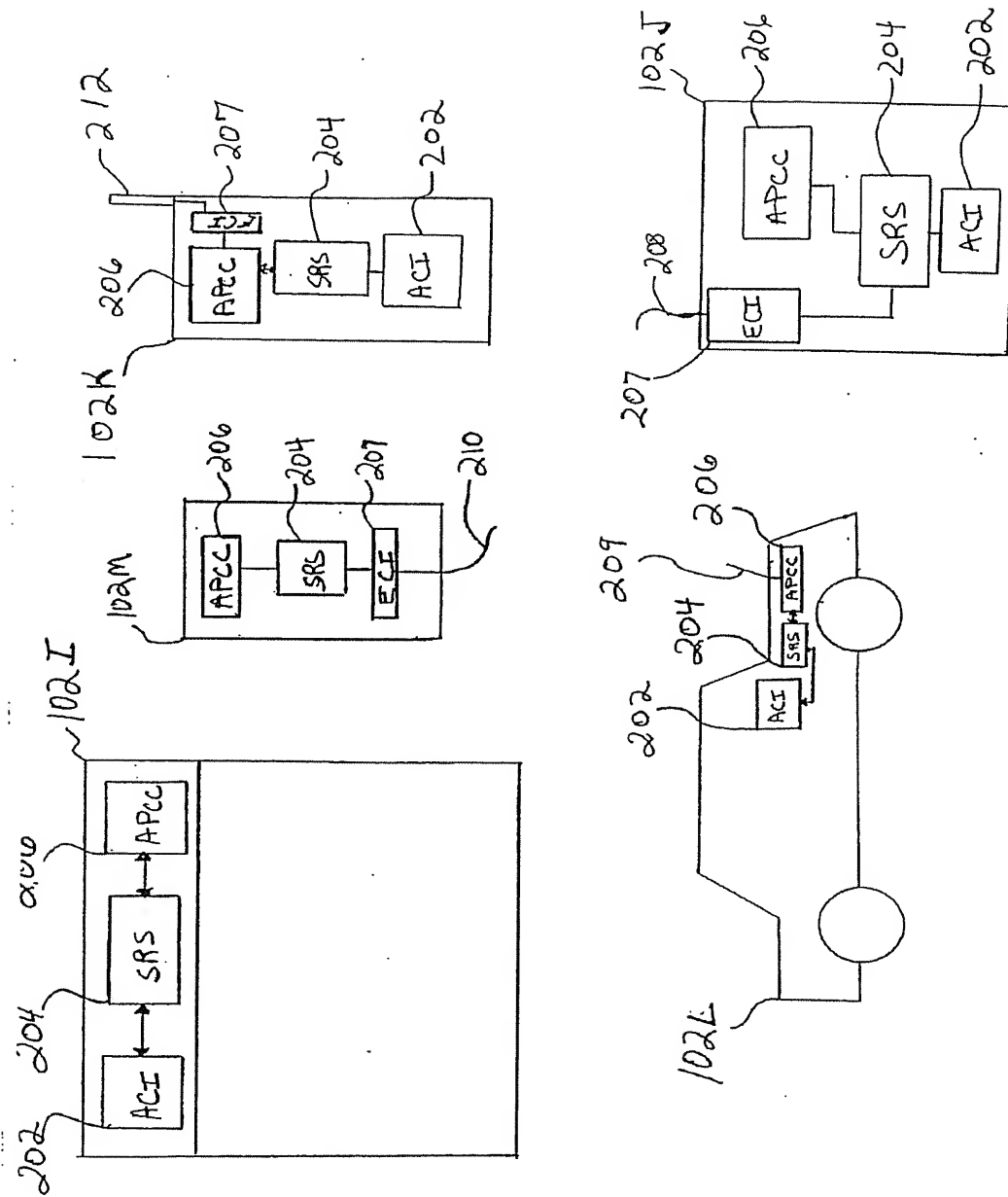


【図1B】

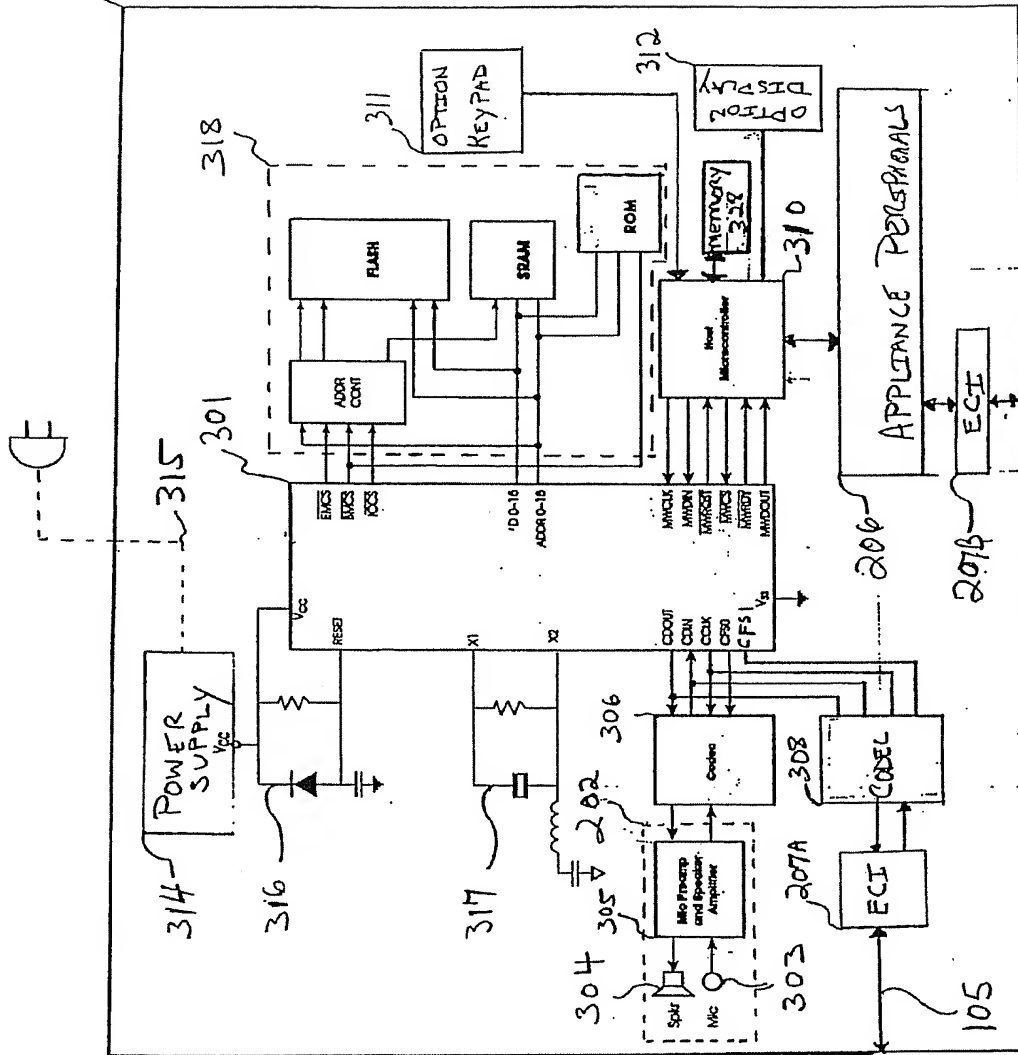


- 112 電話
- 116 通信会社のスイッチ
- 118 コンピュータ

【図2】

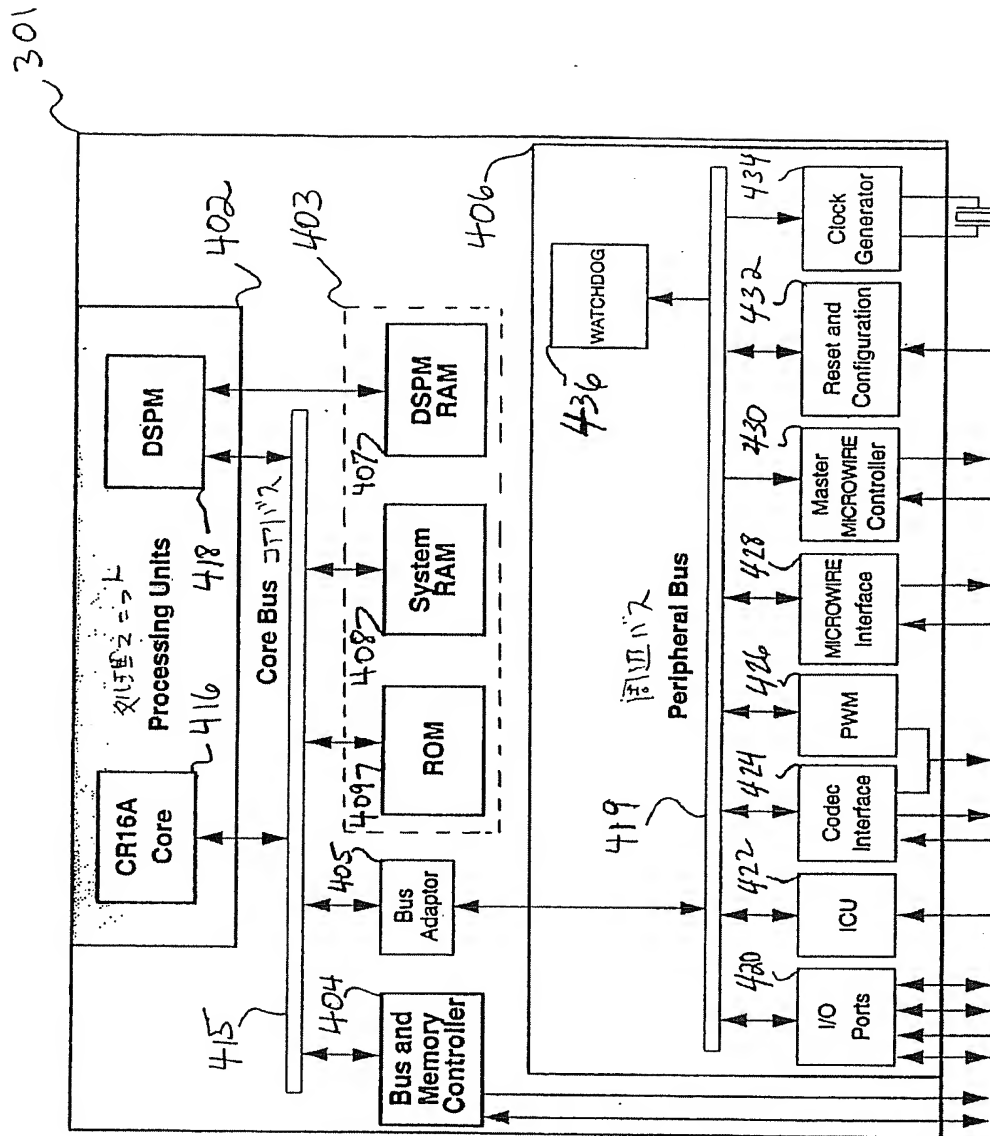


102



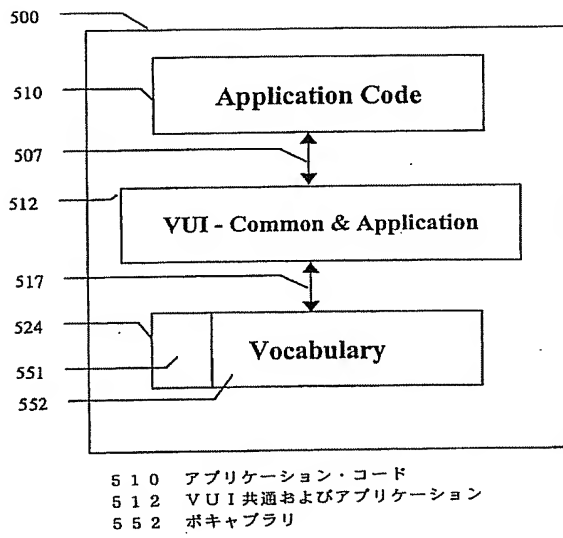
- |     |                    |
|-----|--------------------|
| 206 | 周辺機器               |
| 305 | 前置増幅器と<br>スピーカ増幅器  |
| 306 | コーデック              |
| 310 | ホスト・マイク<br>ロコントローラ |
| 311 | オブションのキ<br>ーパッド    |
| 312 | オブションの<br>ディスプレイ   |
| 314 | 電源                 |
| 328 | メモリ                |

【図4】

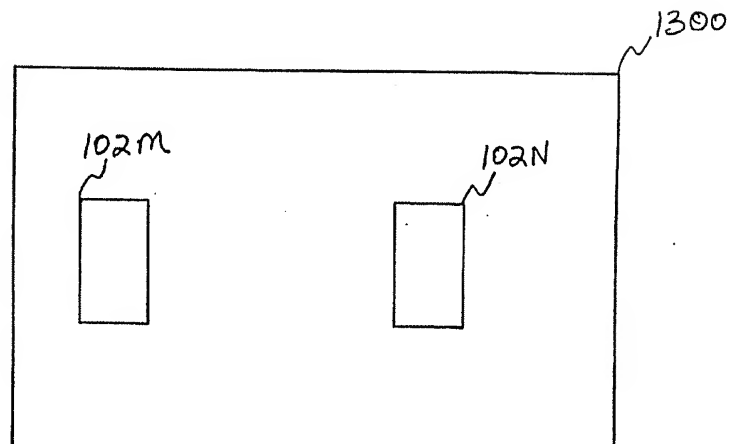


- 404 バスとメモリ  
コントローラ
- 405 バス・アダプタ
- 408 システムRAM
- 420 I/Oポート
- 424 コーデック
- 428 インタフェース  
MICROWIRE
- 430 インタフェース  
マスターMICROWIRE  
コントローラ
- 432 リセットと構成
- 434 クロック生成器
- 436 ウォッチドッグ

【図 5】



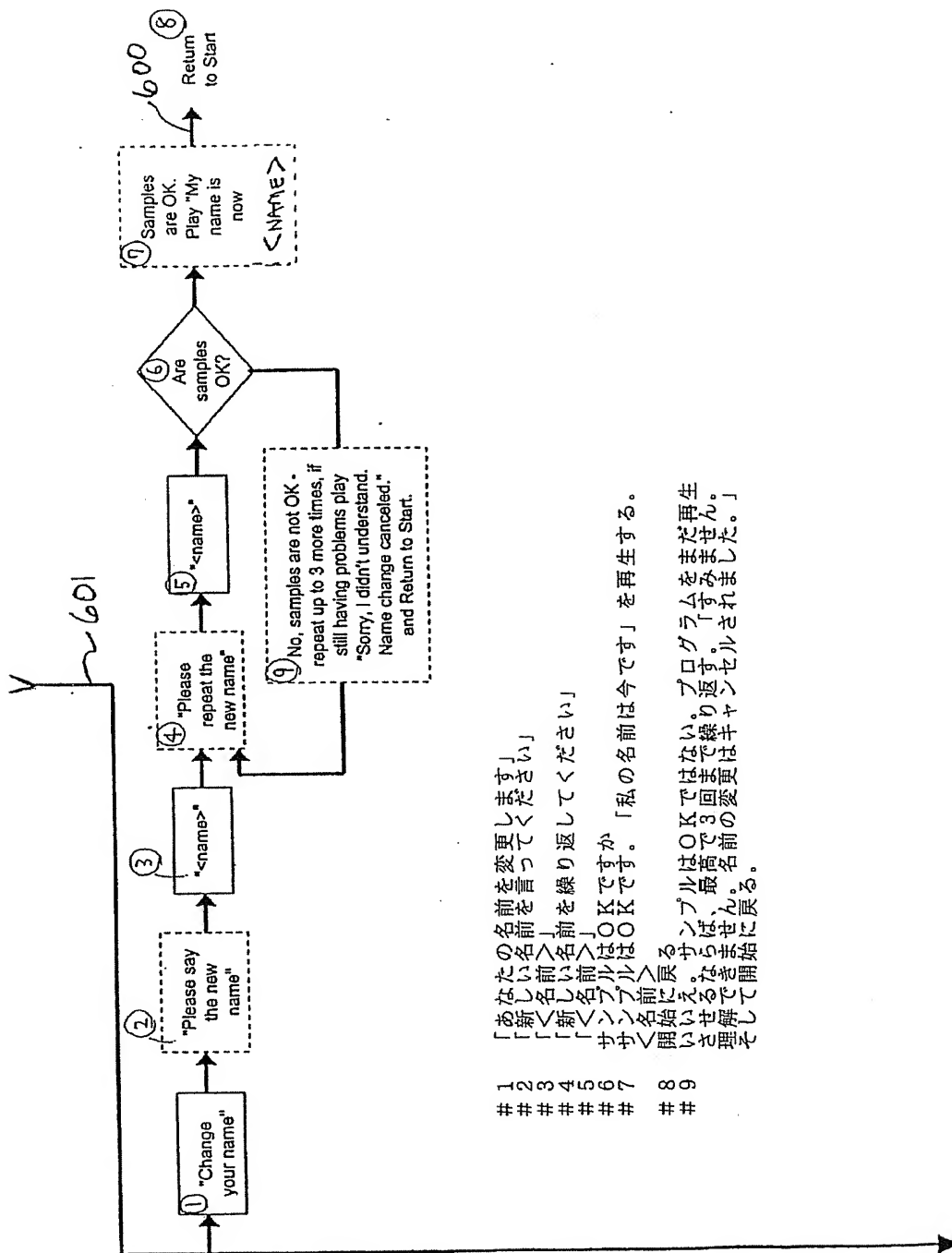
【図 13】



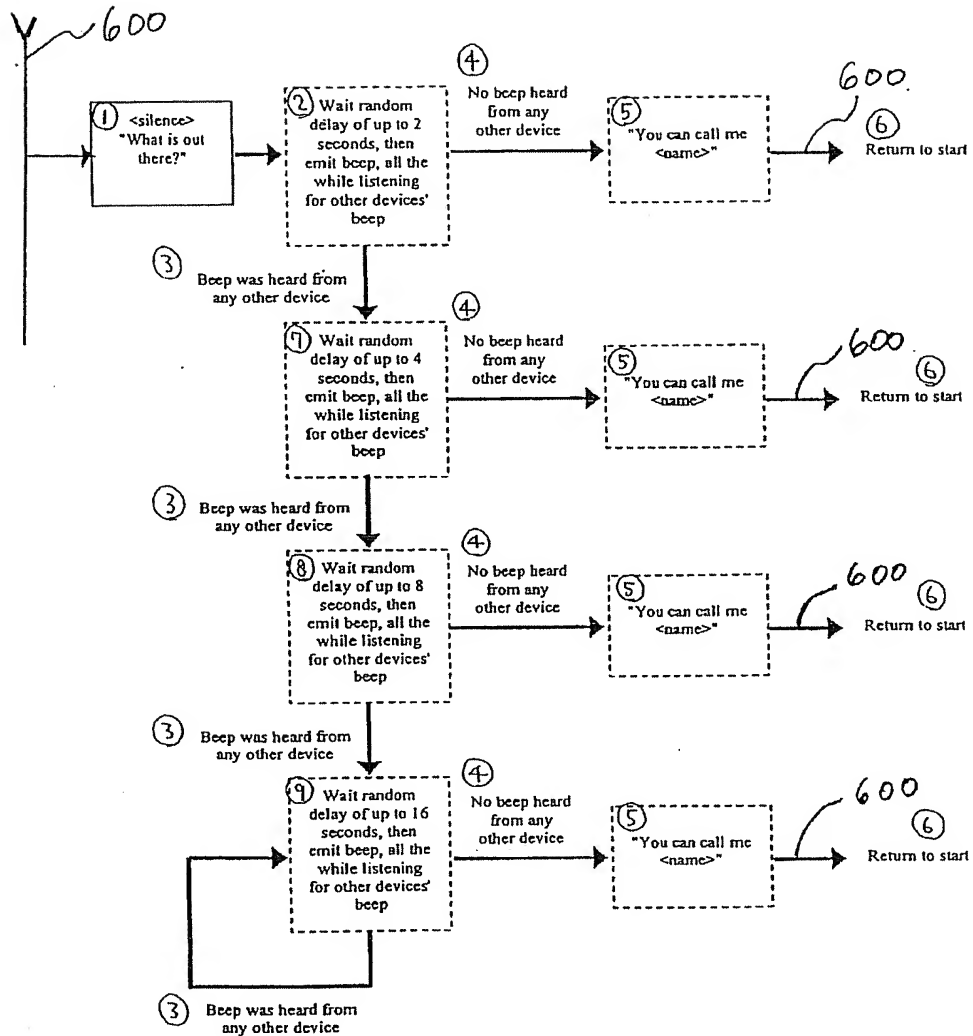


1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9																																								
開	始	触	れ	沈	沈	り	名	言	名	言	名	名	名	6	工	れ	各	開	ン	ボ	「	メ	・	ユ	m	「	セ	ル	再																													
話	す	が	押	さ	れ	る	「	そ	こ	に	何	が	「	あ	だ	な	い	の	グ	を	創	参	照	口	前	に	を	る	際	ミ	ラ	ー	メ	「	一	に	す	」	ス	」	止	ま	」	停	る	再												
始	触	れ	沈	沈	り	名	言	名	言	名	名	名	名	6	工	れ	各	開	ン	ボ	「	メ	・	ユ	m	「	セ	ル	再	始	触	れ	沈	沈	り	名	言	名	言	名	名	名	6	工	れ	各	開	ン	ボ	「	メ	・	ユ	m	「	セ	ル	再

【図 6 B】



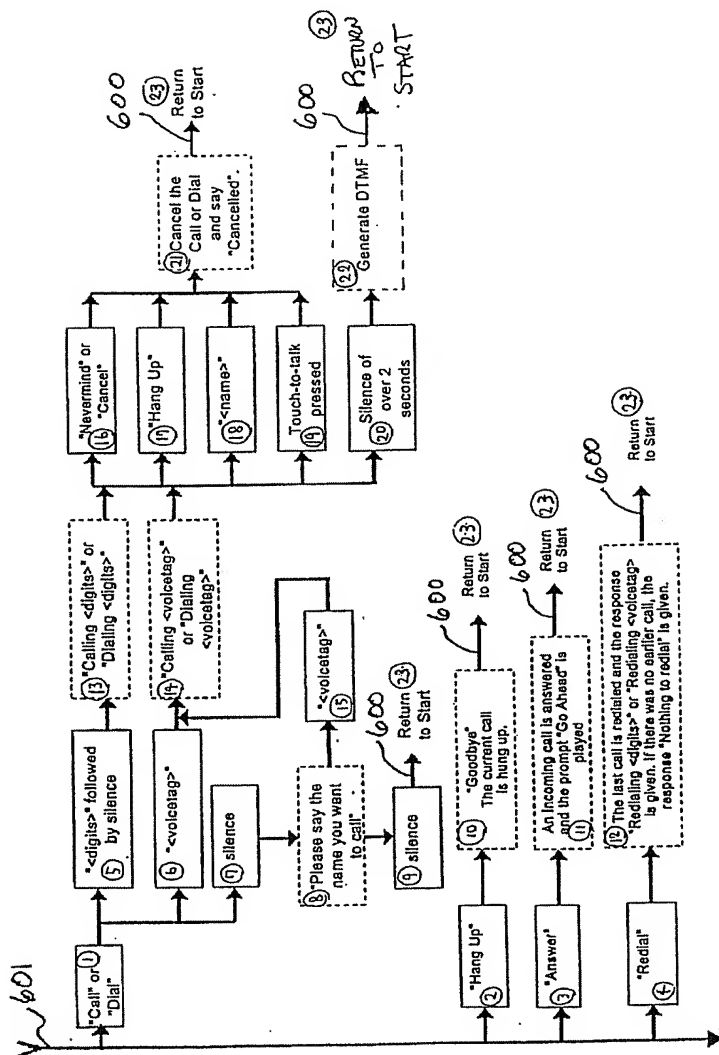
【図 6 C】



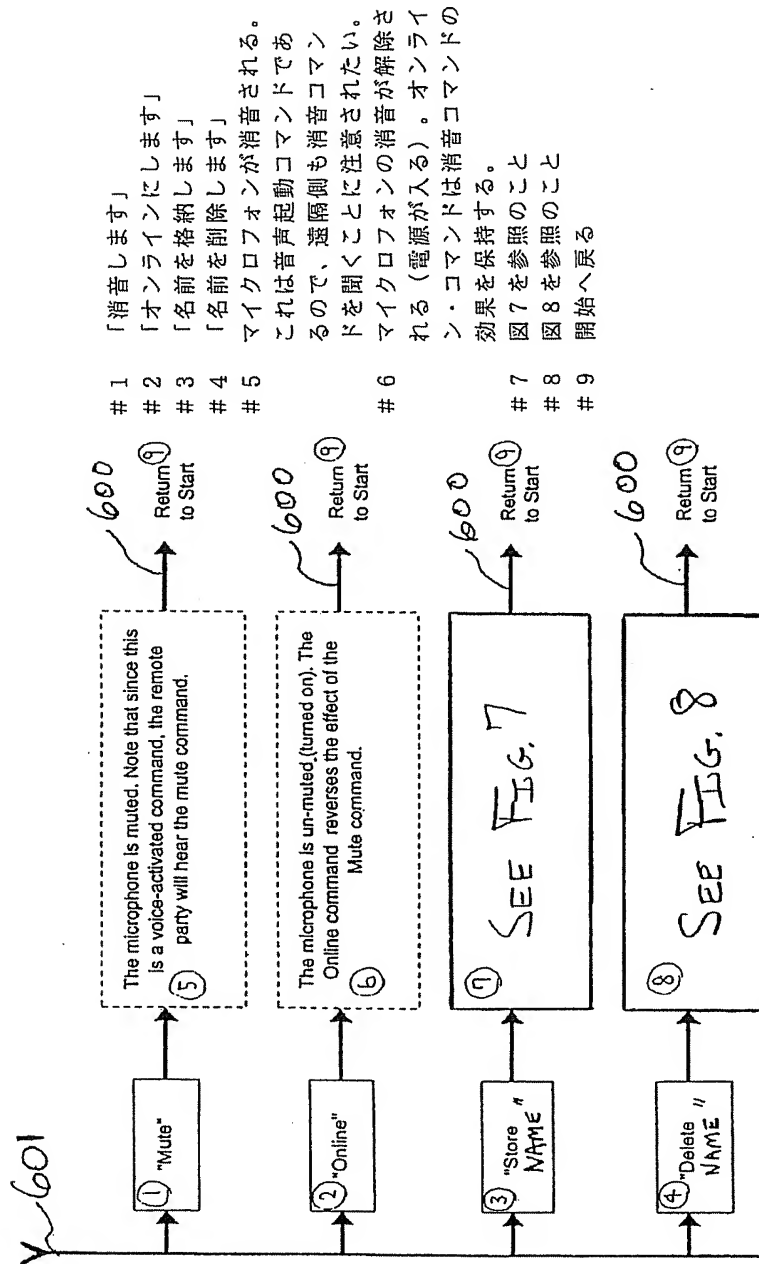
# 1 <沈黙> 「そこを待たせよう、名前を聞いてみる」  
 # 2 他装置の遅延を待たせよう、名前を聞いてみる  
 # 3 ランダムに遅延を待たせよう、名前を聞いてみる  
 # 4 任意の遅延を待たせよう、名前を聞いてみる  
 # 5 開始の遅延を待たせよう、名前を聞いてみる  
 # 6 他装置の遅延を待たせよう、名前を聞いてみる  
 # 7 ランダムに遅延を待たせよう、名前を聞いてみる  
 # 8 他装置の遅延を待たせよう、名前を聞いてみる  
 # 9 ランダムに遅延を待たせよう、名前を聞いてみる

中、最高で 2 秒までの  
 間、名前を聞いてみる  
 中、最高で 4 秒までの  
 間、名前を聞いてみる  
 中、最高で 8 秒までの  
 間、名前を聞いてみる  
 中、最高で 16 秒までの  
 間、名前を聞いてみる

【図 6 D】

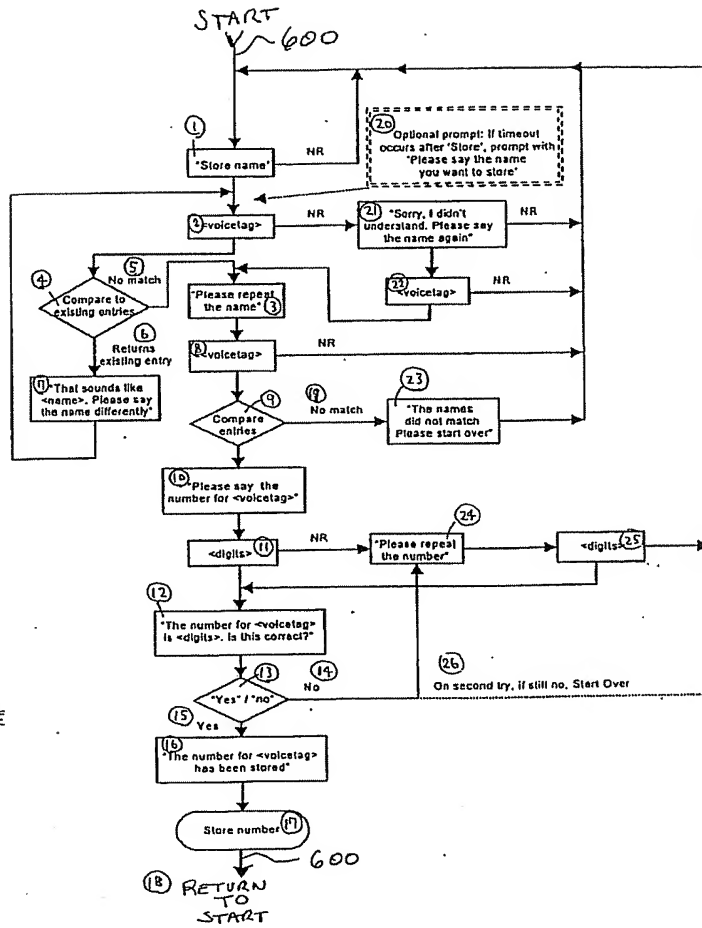
[illegible]

【図 6 E】

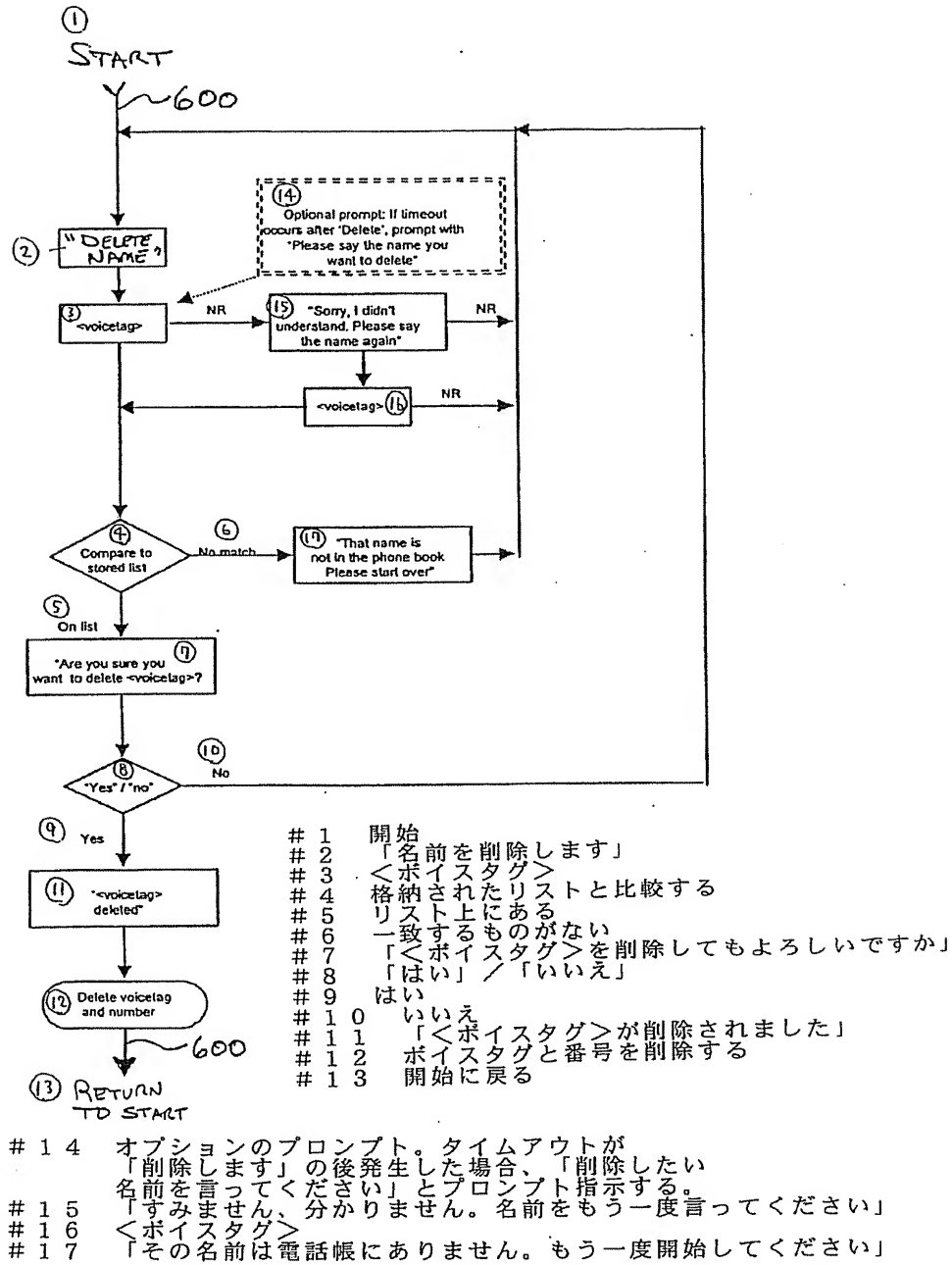


# 1 「消音します」  
 # 2 「オンラインにします」  
 # 3 「名前を格納します」  
 # 4 「名前を削除します」  
 # 5 マイクロフォンが消音される。  
 これは音声起動コマンドであるので、遠隔側も消音コマンドを聞くことに注意されたい。  
 # 6 マイクロフォンの消音が解除される（電源が入る）。オンライン・コマンドは消音コマンドの効果を持続する。  
 # 7 図 7 を参照のこと  
 # 8 図 8 を参照のこと  
 # 9 開始へ戻る

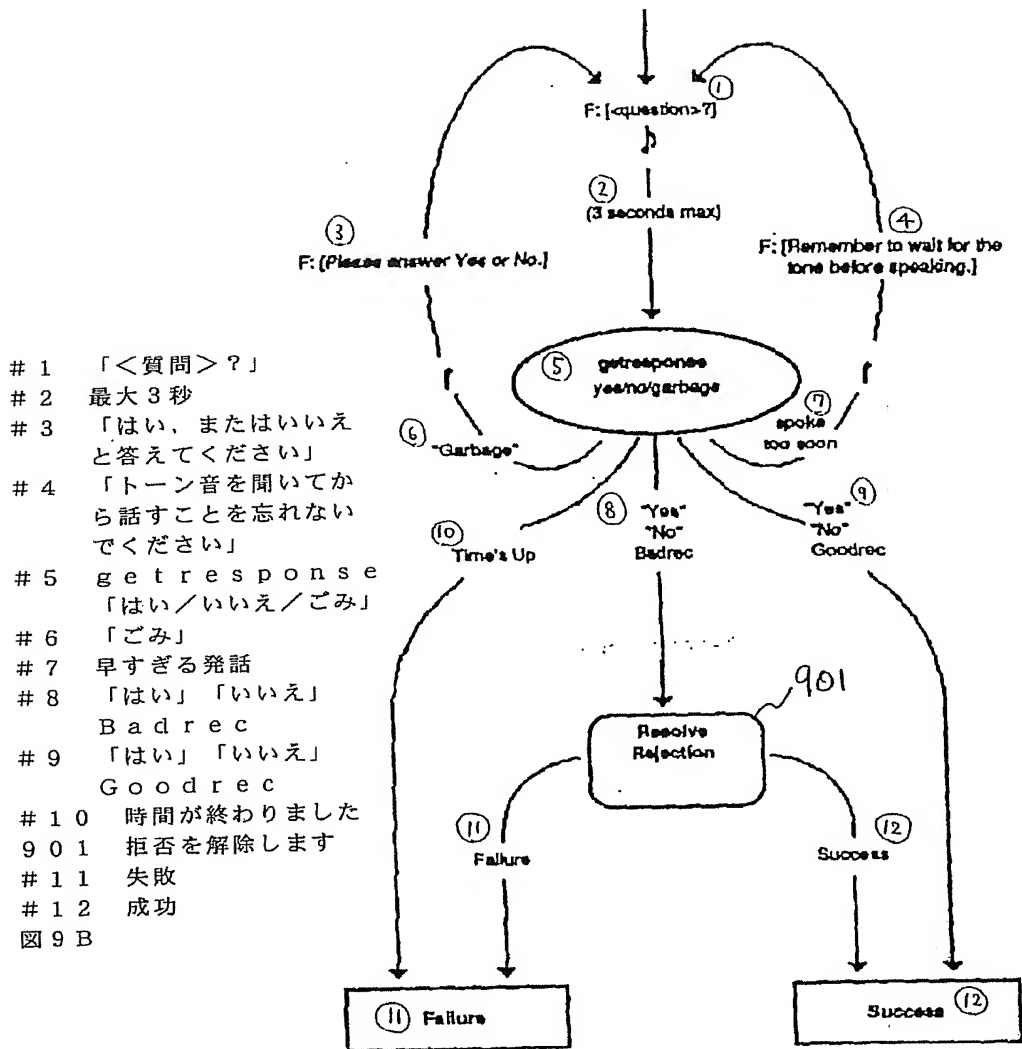
【図 7】

[illegible]

【图 8】

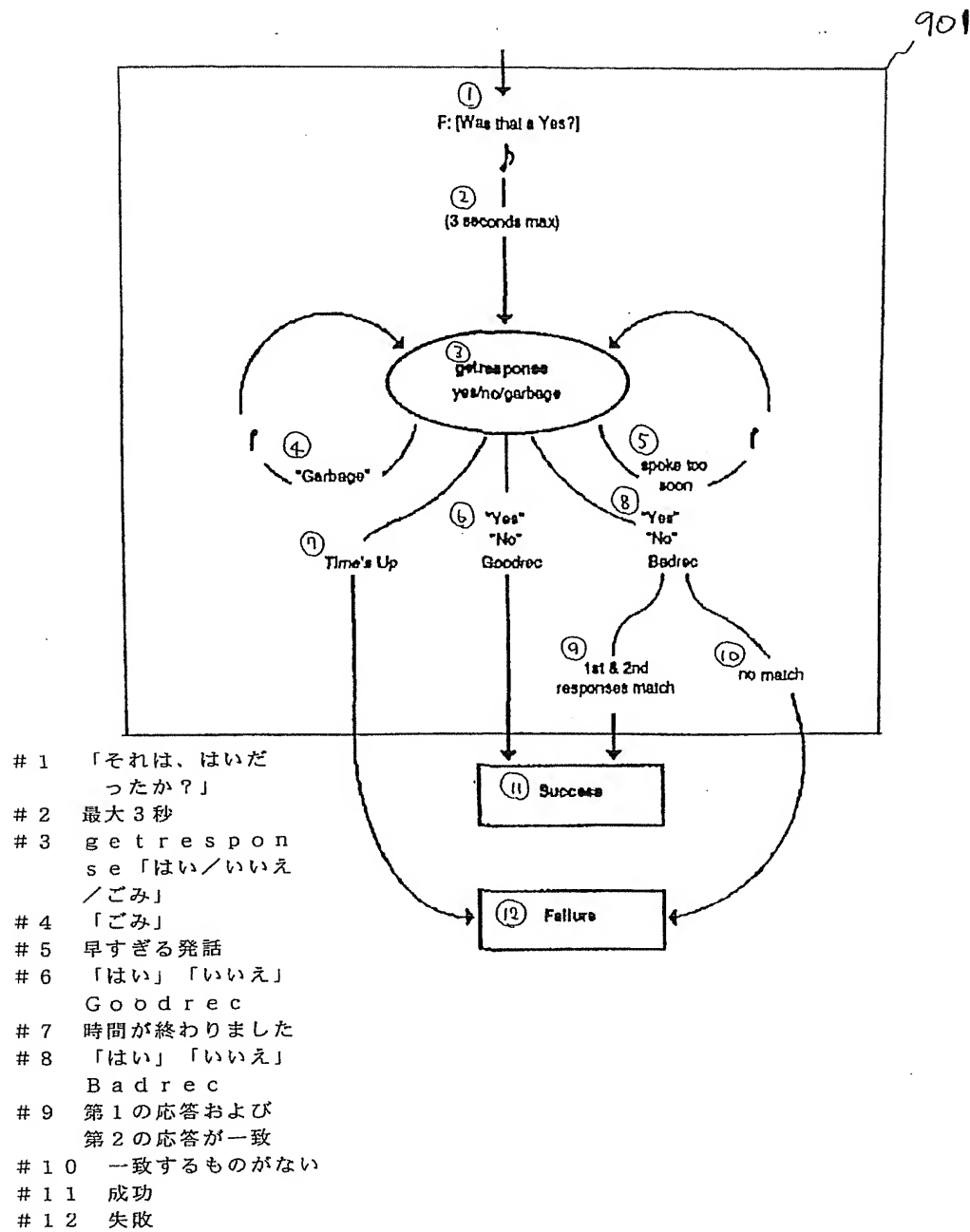


【図 9 A】

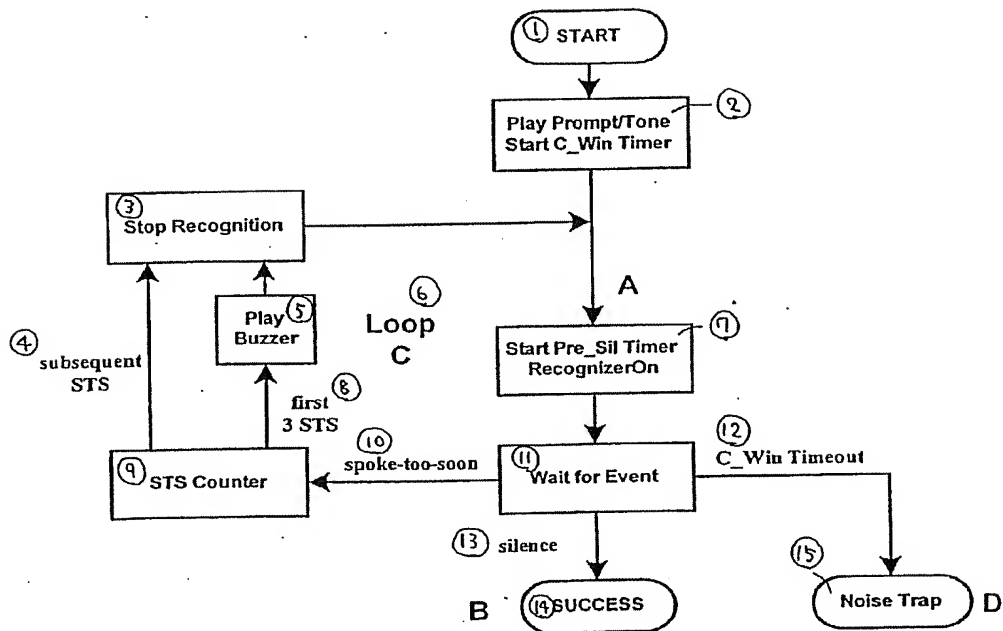




【図9B】

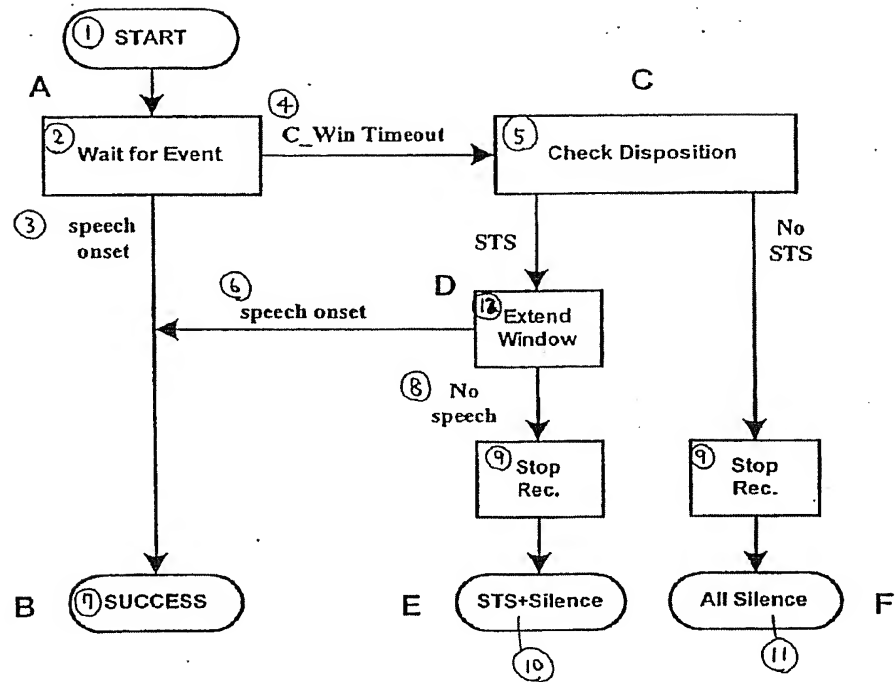


【図 10 A】



# 1 開始  
 # 2 プロンプト/トーンの再生  
 # 3 C\_Winタイマの開始  
 # 4 認識の停止 STS  
 # 5 プレイヤーの再生  
 # 6 プレイヤーの再生  
 # 7 Pre\_Sil タイマ・認識器起動の開始  
 # 8 first 3 STS  
 # 9 STS  
 # 10 spoke-too-soon  
 # 11 Wait for Event  
 # 12 C\_Win タイムアウト  
 # 13 silence  
 # 14 成功  
 # 15 ノイズ・トラップ

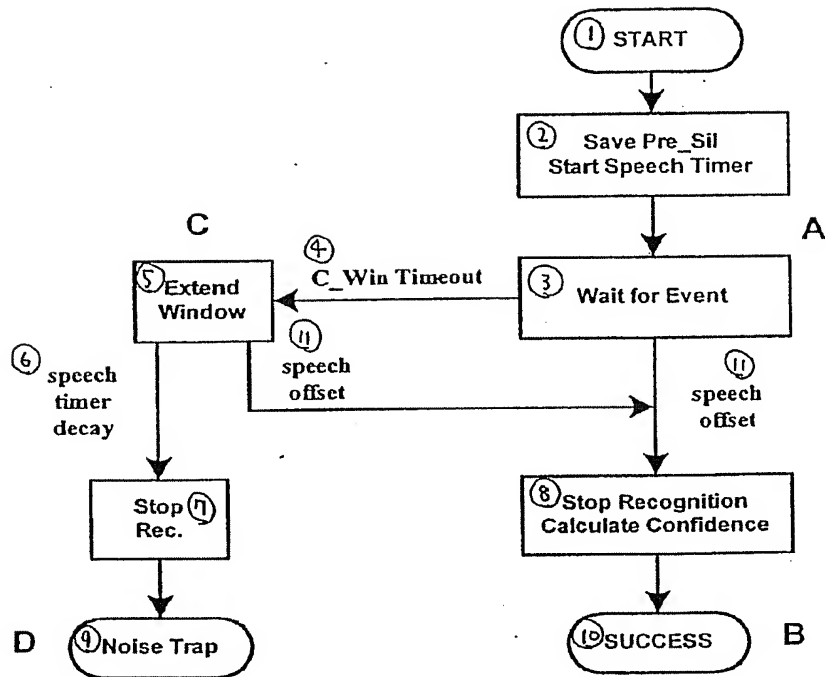
【図 10 B】



# 1 開始  
 # 2 イベントを待つ  
 # 3 発話オンセット  
 # 4 C\_Winタイムアウト  
 # 5 チェックの配置  
 # 6 発話のオンセット

# 7 成功  
 # 8 発話なし  
 # 9 記録の停止  
 # 10 STS+沈黙  
 # 11 すべての沈黙を拡張する  
 # 12 ウィンドウを拡張する

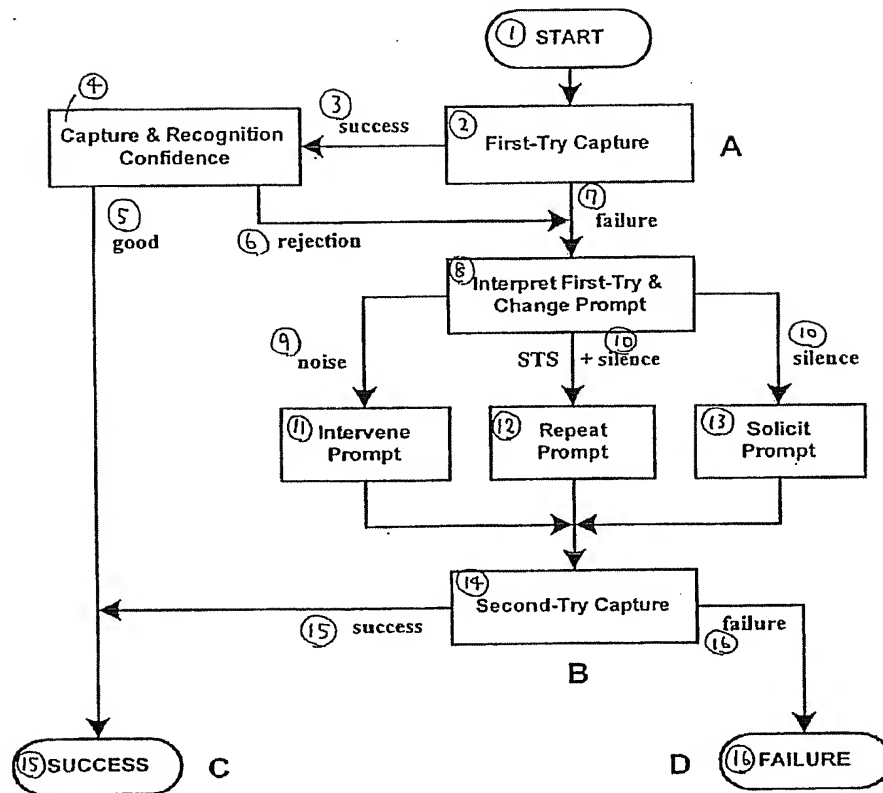
【図10C】



# 1 開始  
 # 2 Pre\_Silの保存  
 # 3 イベントの開始  
 # 4 C\_Win Timeout  
 # 5 Extend Window  
 # 6 speech timer decay  
 # 7 Stop Rec.  
 # 8 Stop Recognition Calculate Confidence  
 # 9 Noise Trap  
 # 10 SUCCESS

# 7 記録の停止  
 # 8 認識の停止  
 # 9 ノイズトラップ  
 # 10 成功  
 # 11 発話オフセット

【図 11】



# 1 開始  
 # 2 第 1 の試み捕捉  
 # 3 成功捕捉と認識、信頼性  
 # 4 捕捉と認識、信頼性  
 # 5 拒否  
 # 6 拒否  
 # 7 失敗

# 8 第 1 の試みの解釈と、プロンプトの変更  
 # 9 雑音  
 # 10 沈黙  
 # 11 介入プロンプト  
 # 12 プロンプトを繰り返す  
 # 13 プロンプトを促す  
 # 14 第 2 の試みの捕捉  
 # 15 成功  
 # 16 失敗

```

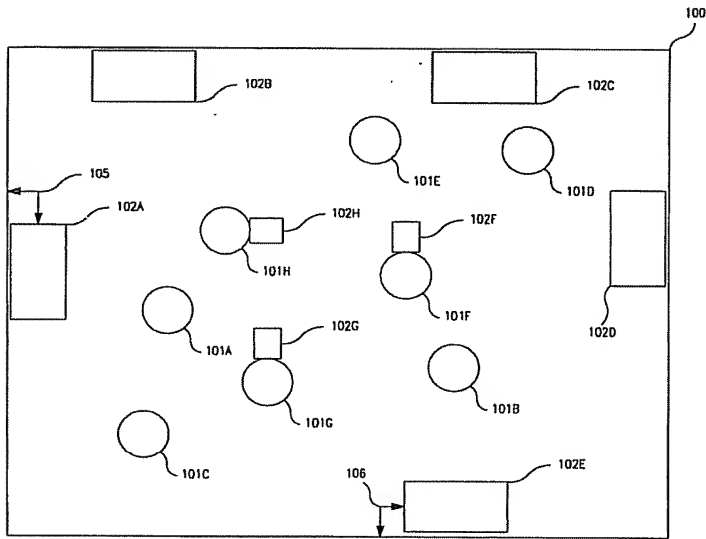
graph TD
    Start(( )) --> J14((14))
    J14 --> J15{15  
rec  
good  
?}
    J15 -- Yes 11 --> J17{17  
this  
item  
?}
    J15 -- No 10 --> J19{19  
Yes/No  
[this item]}
    J17 -- Yes 11 --> J22{22  
Menu Item  
Selected}
    J17 -- No 10 --> J18{18  
Yes/No  
[rec item]}
    J18 -- Yes 11 --> J21{21  
Recognized  
Item Selected}
    J18 -- No 10 --> J19
    J19 -- Yes 11 --> J19
    J19 -- No 10 --> J16{16  
spurious  
response}
    J16 --> J13{13  
word}
    J13 --> J15
    J21 --> J23((23  
Success))
    J22 --> J23
    J19 --> J23

```

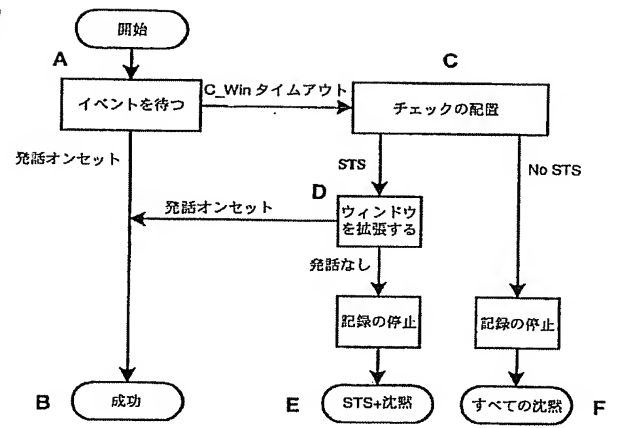
選單照記にははこ選選成  
 折語会録せのいいの折折功  
 なし  
 とはの／アささ  
 選正応いいいれ  
 論いムええム  
 理かか（記こ選識ユ  
 ムム）  
 テテたアテ  
 イイれたイ  
 アアされア  
 録の折さ一  
 ムム）  
 テム

【補正内容】

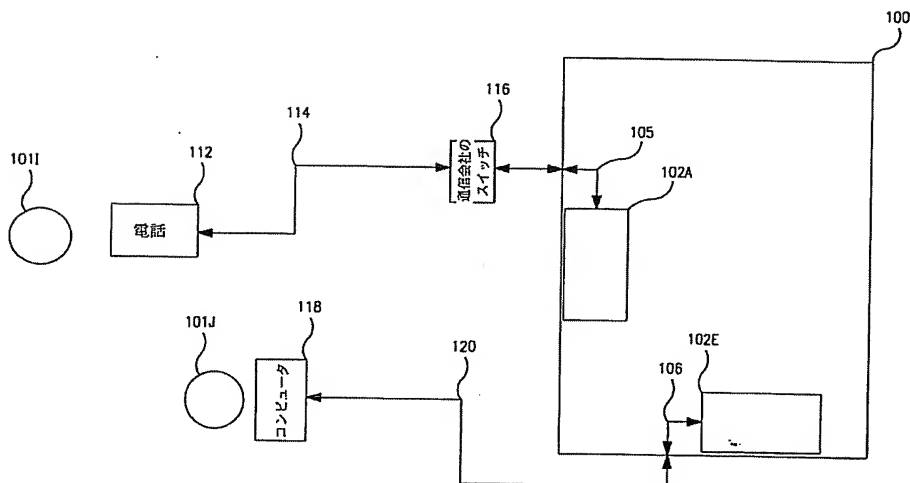
【図 1 A】



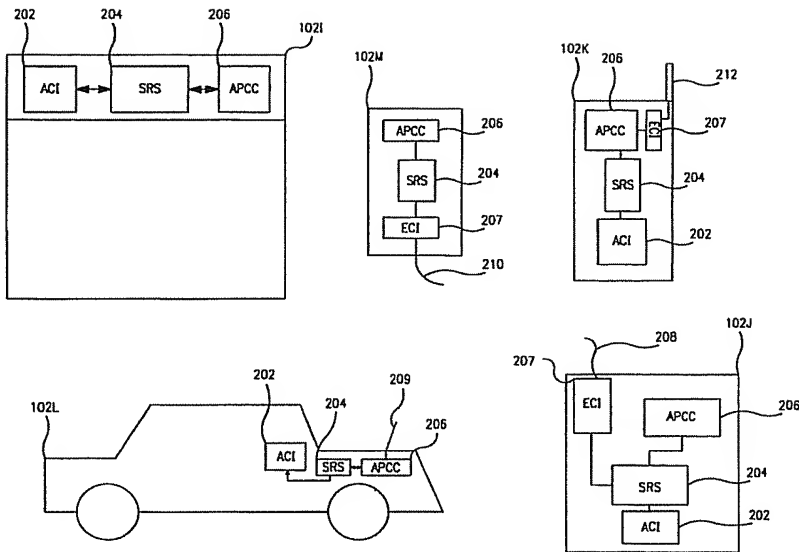
【図 10 B】



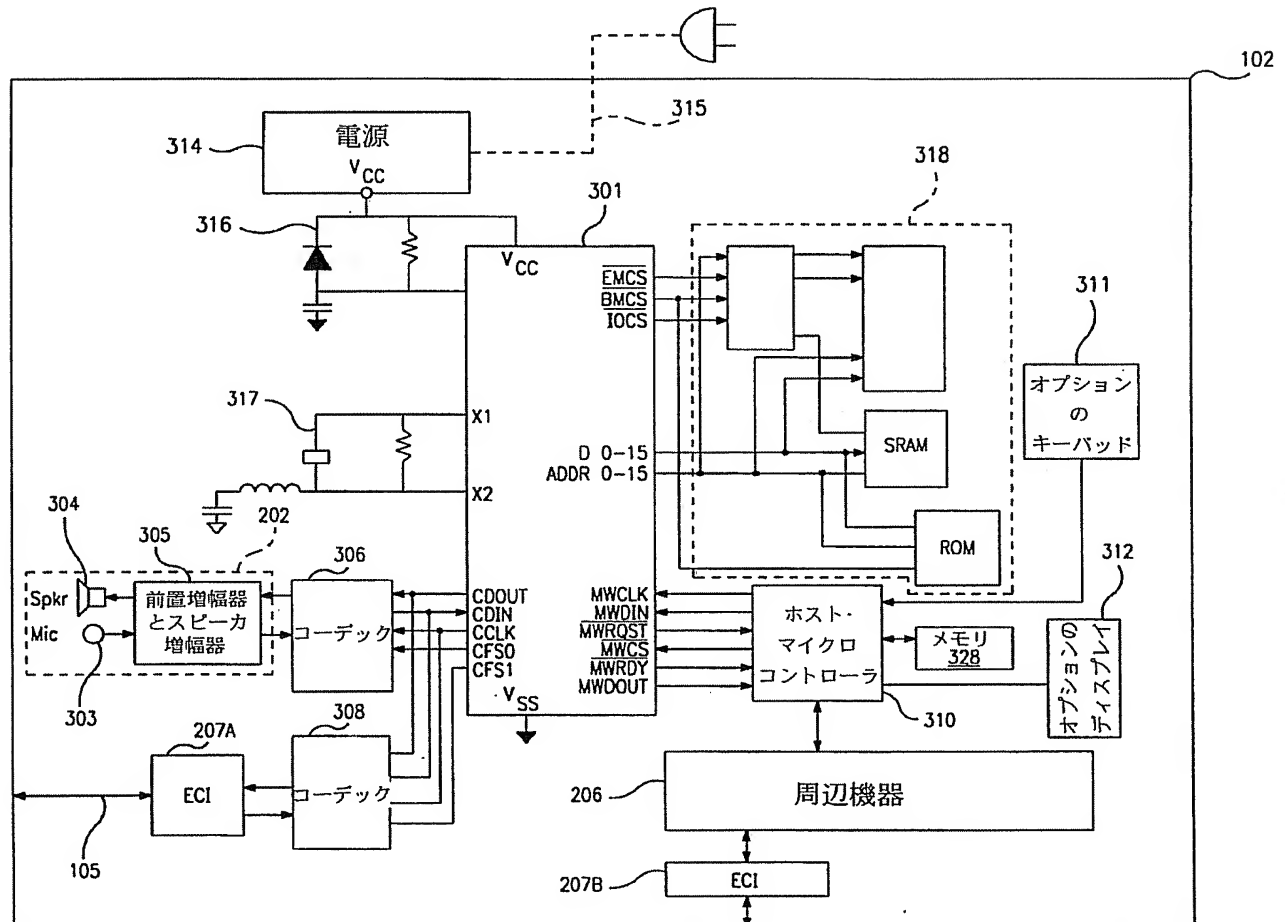
【図 1 B】



【図 2】

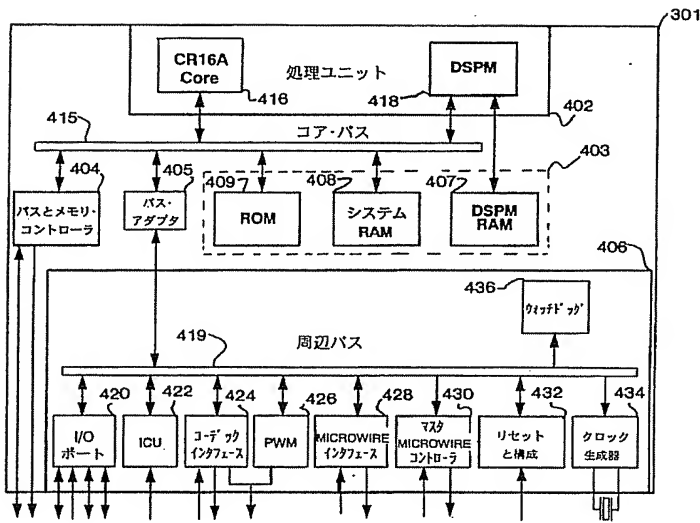


【図 3】

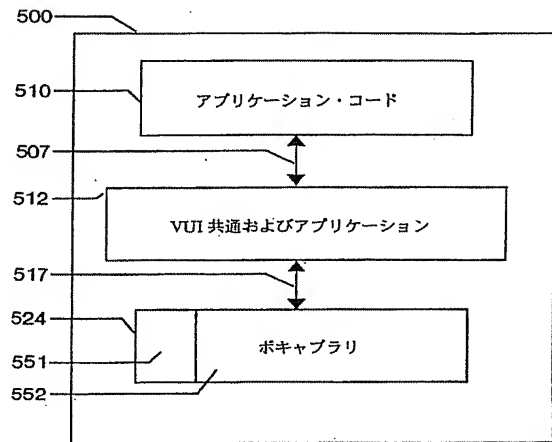




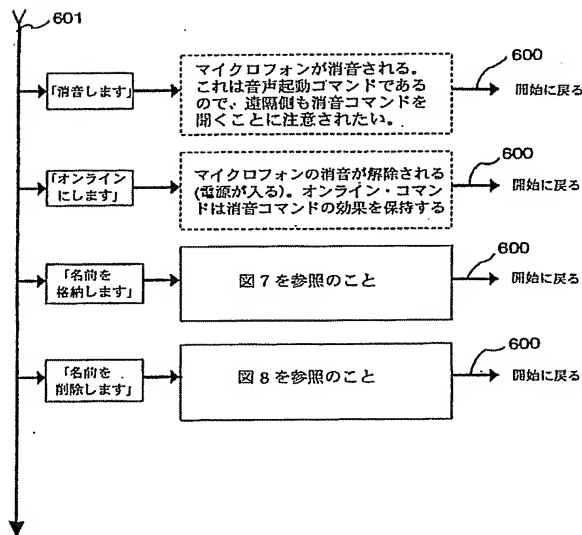
【図 4】



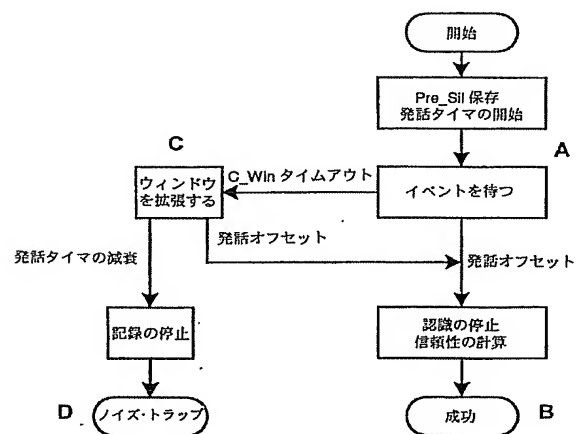
【図 5】



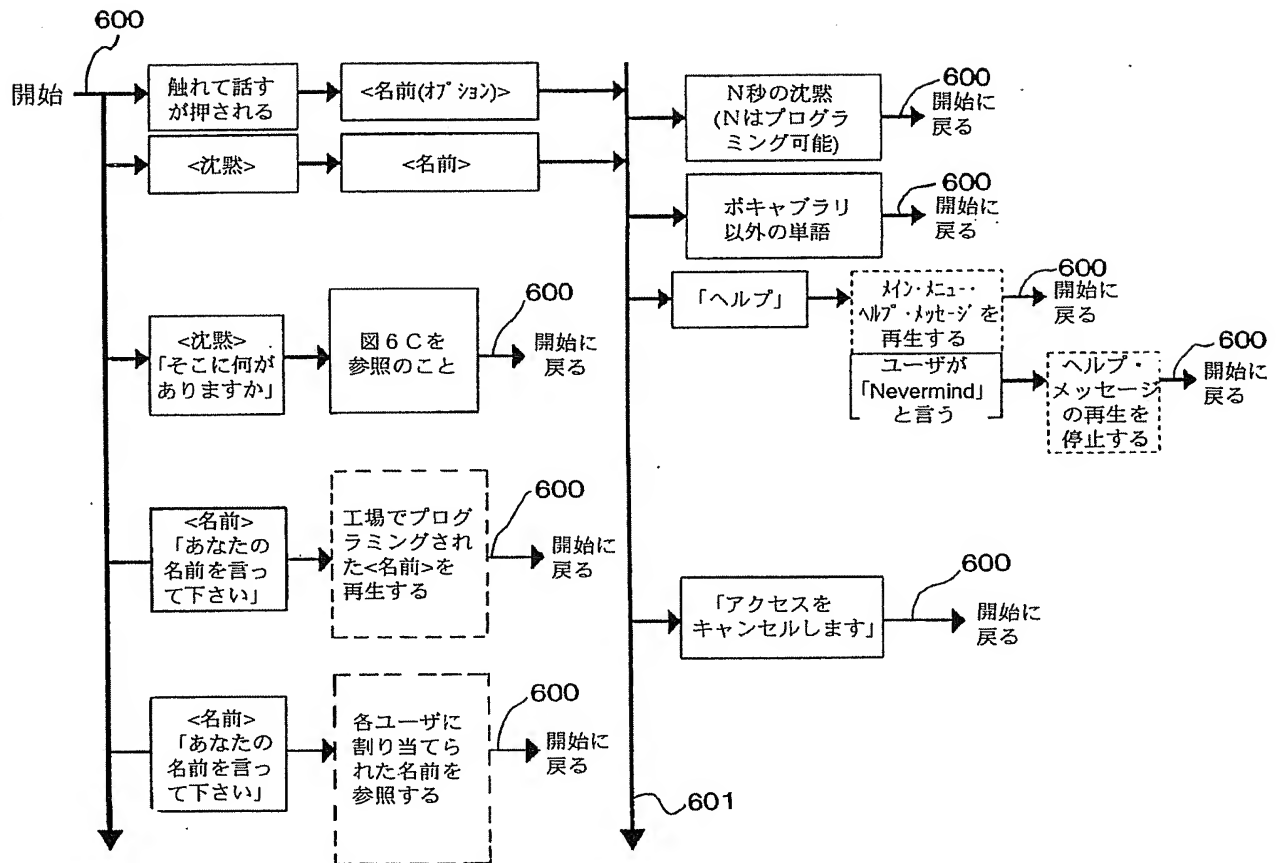
【図 6 E】



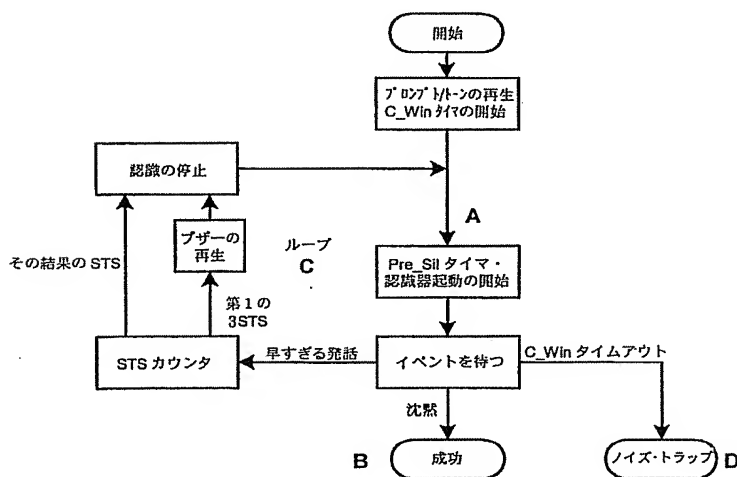
【図 10 C】



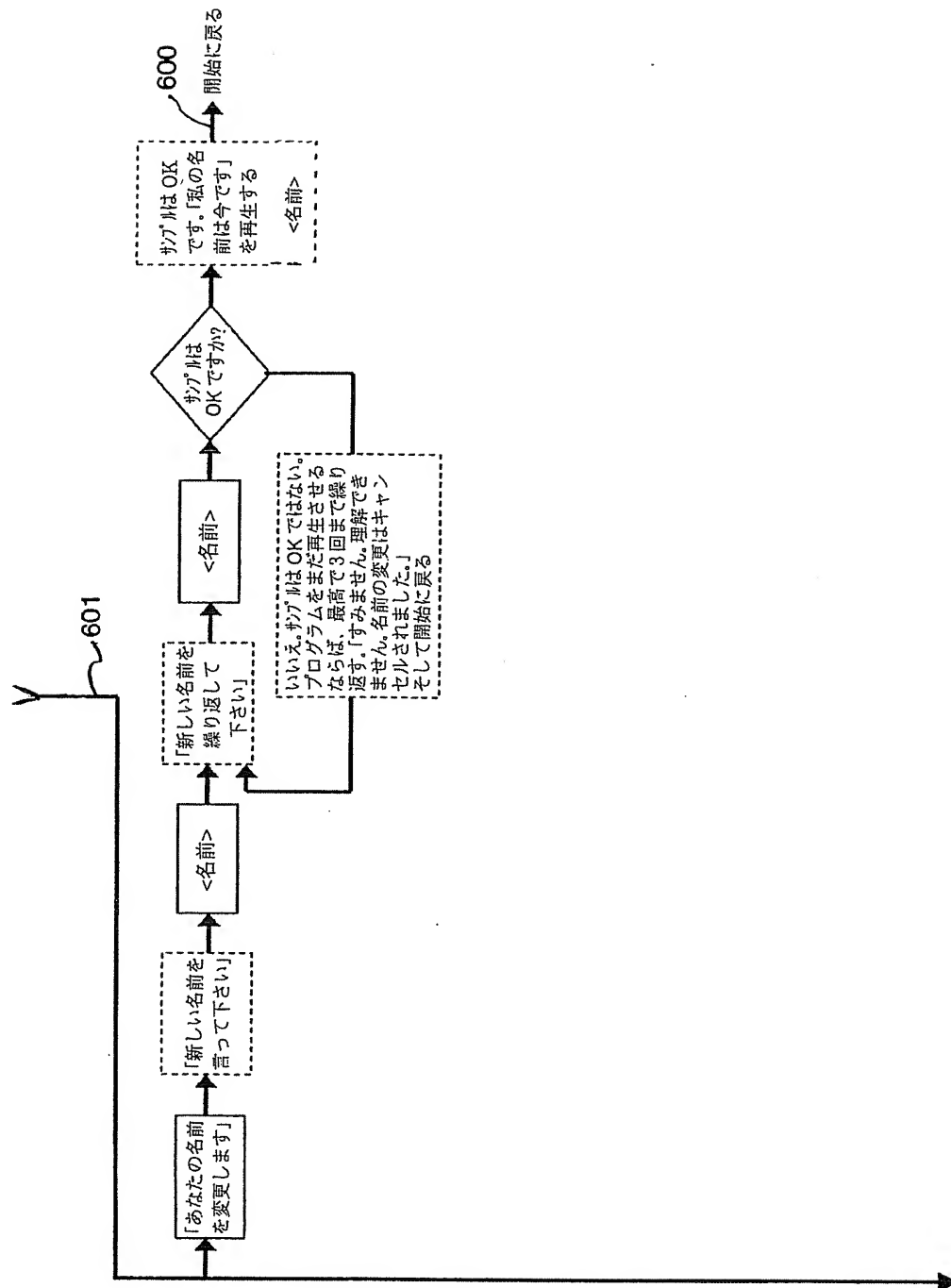
【図 6 A】



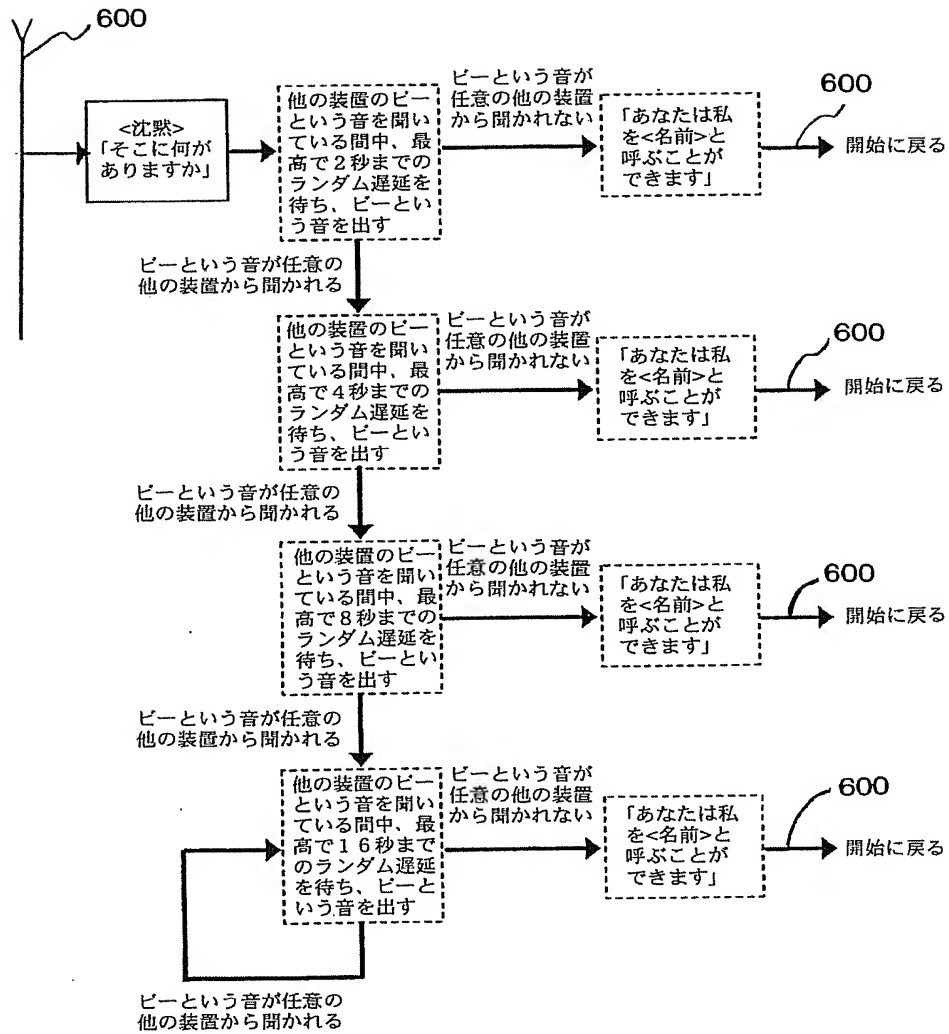
【図 10 A】



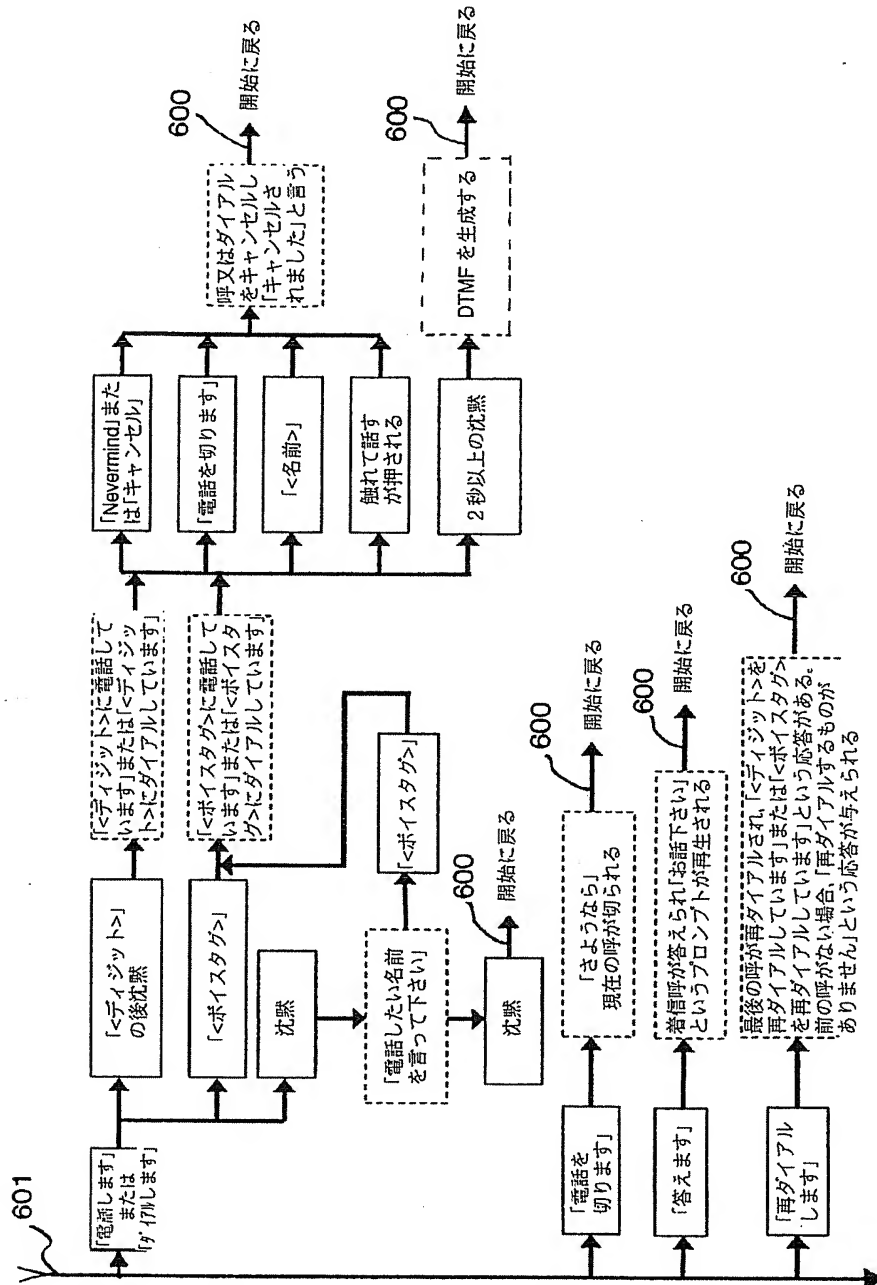
【図6B】



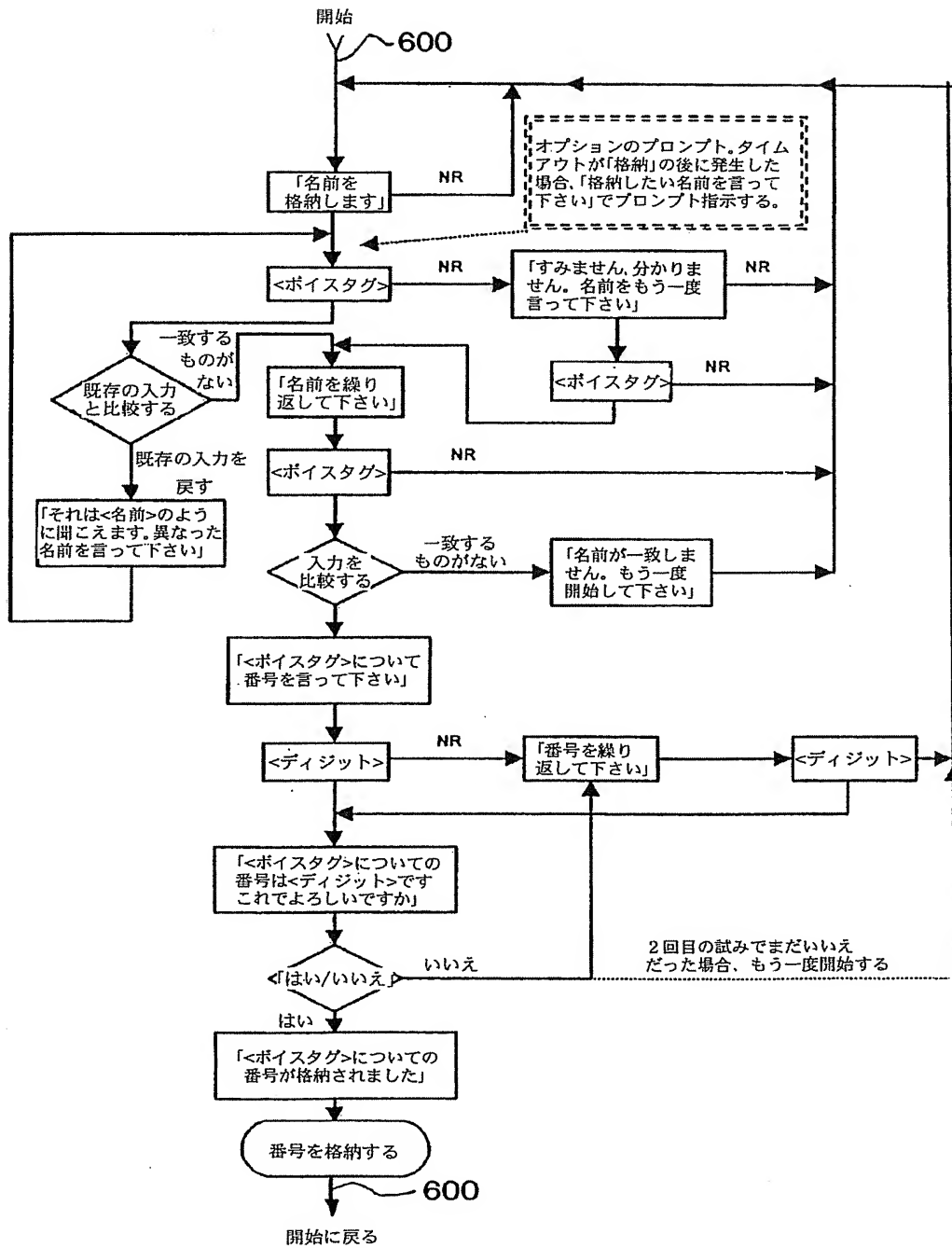
【図 6 C】



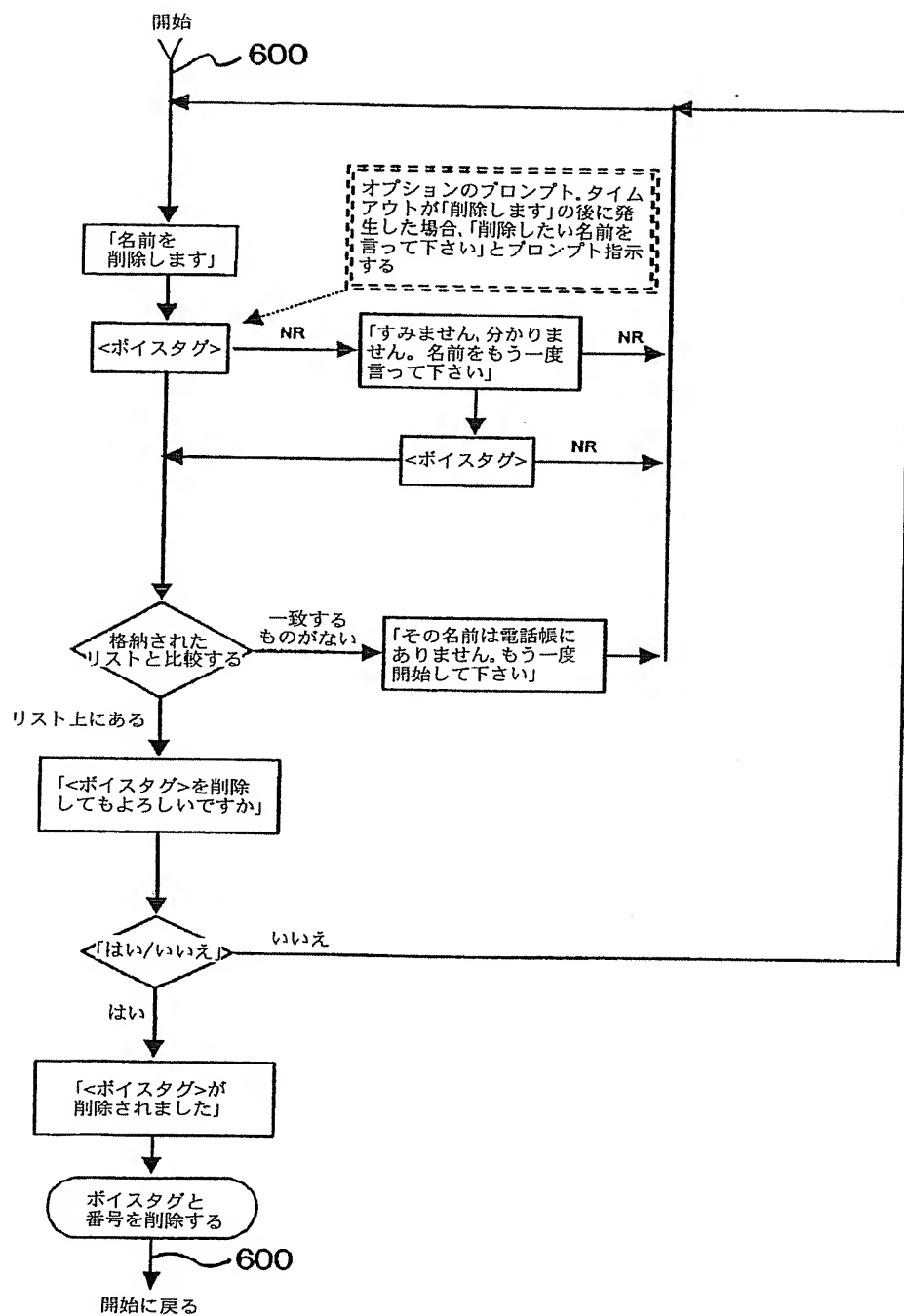
20



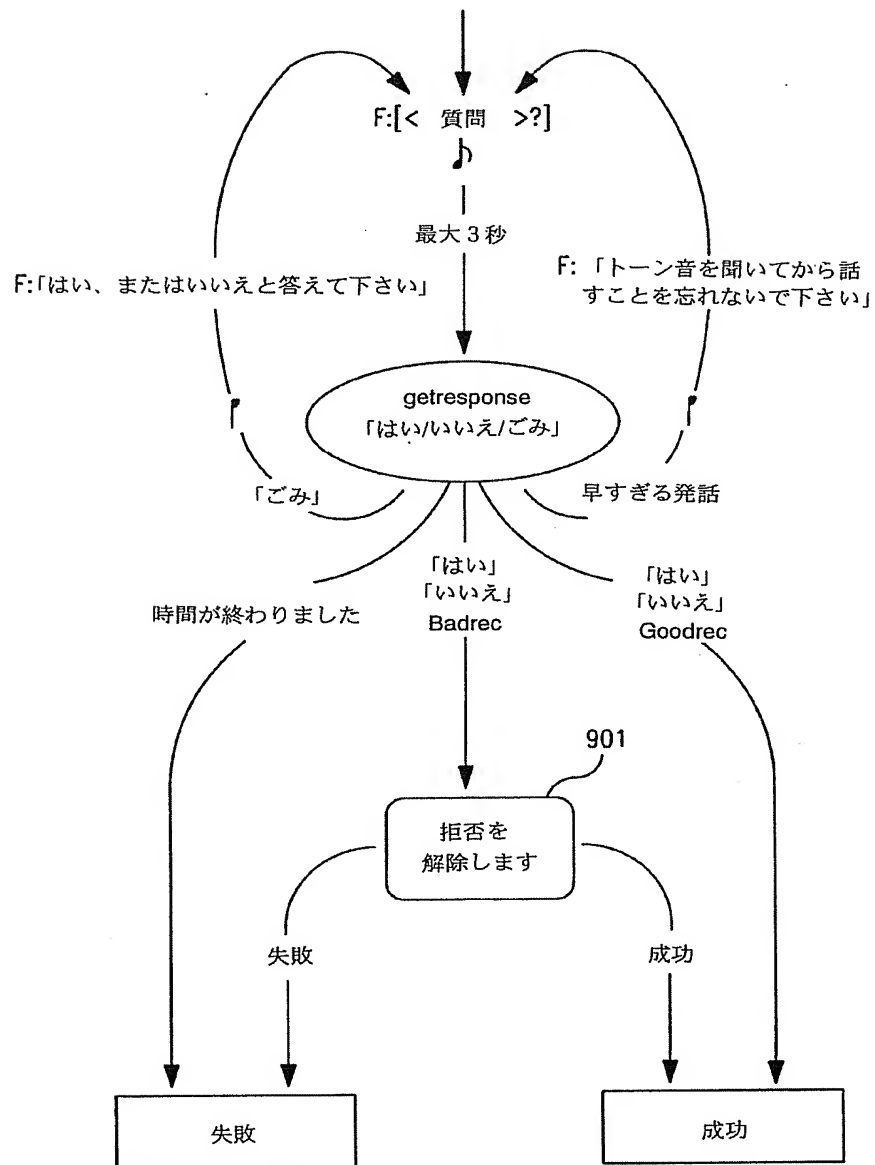
【図7】



【図8】

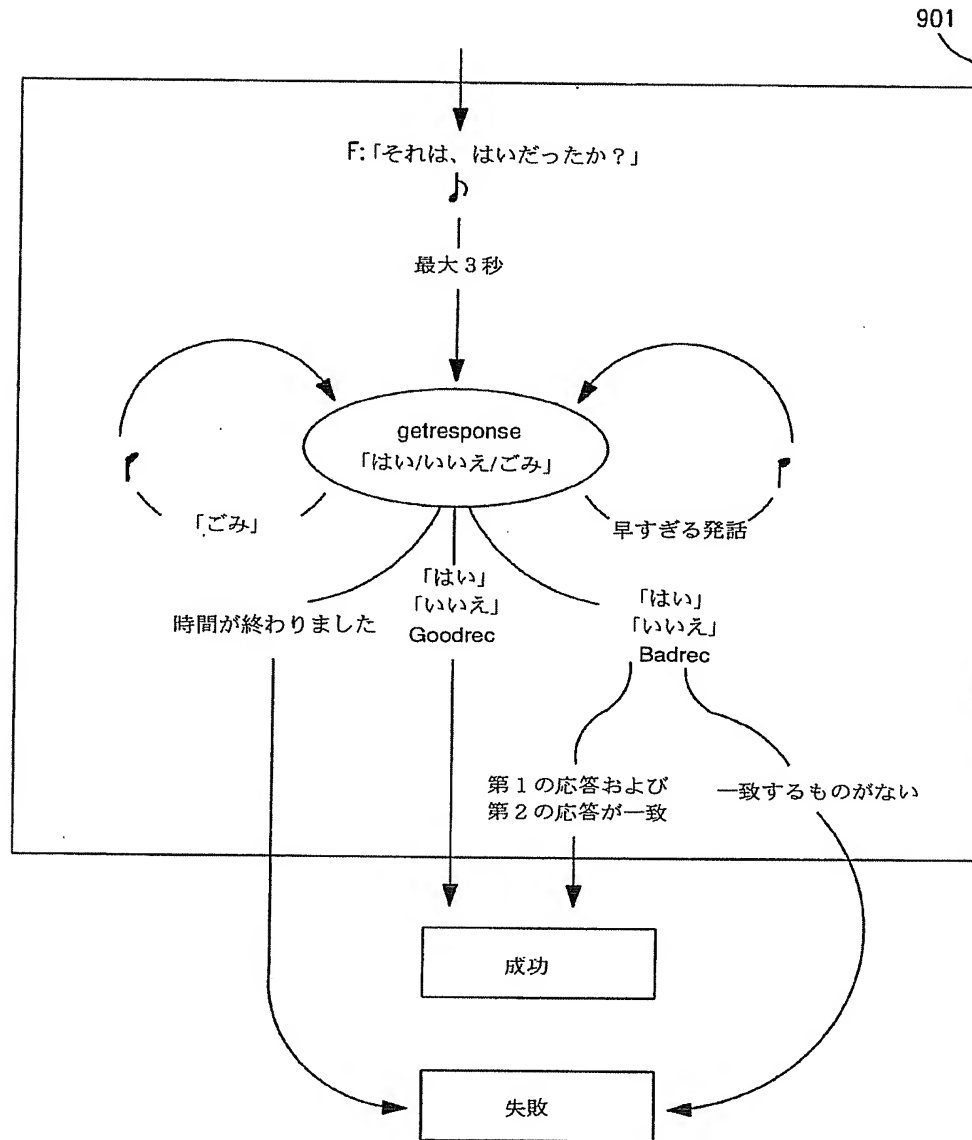


【図 9 A】

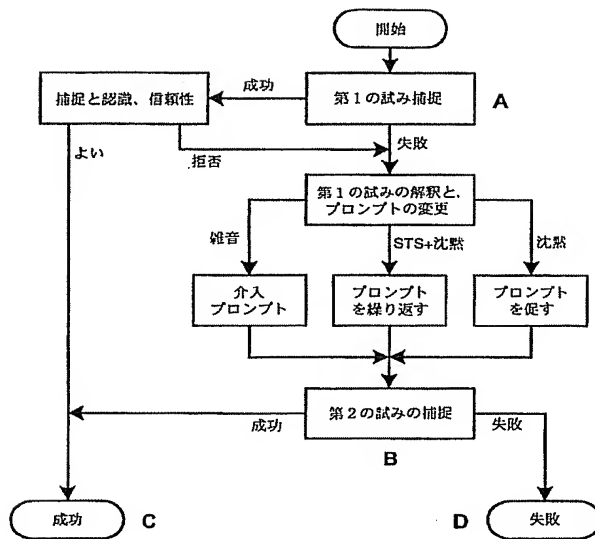




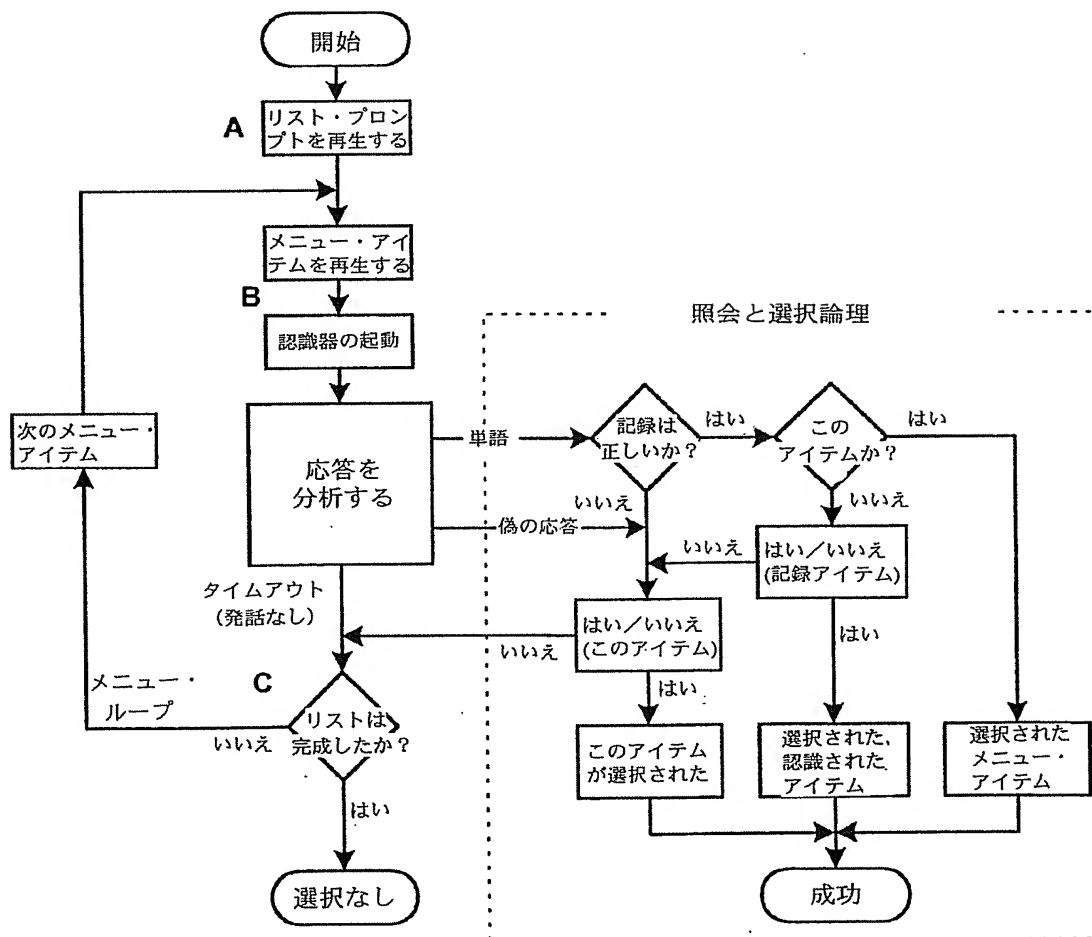
【図9B】



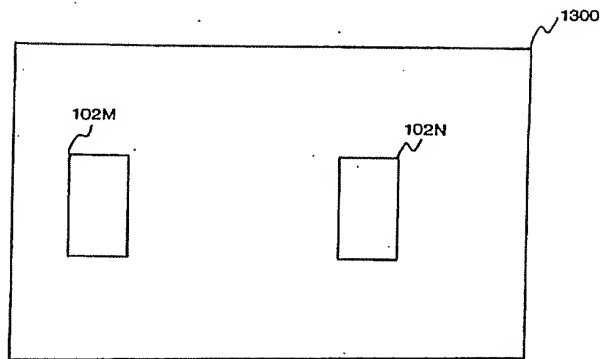
【図11】



【図12】



【図 13】



フロントページの続き

(51) Int. Cl. <sup>7</sup>	識別記号	FI	テーマコード (参考)
G10L 15/22		G10L 3/00	561H 571K 571H 571V
(72) 発明者 デイビッド・マックミラン アメリカ合衆国・94062・カリフォルニア 州・ウッドサイド・ジェファーソン アベ ニュー・3931		(72) 発明者 カリン・リセット・ブーツマ アメリカ合衆国・95134・カリフォルニア 州・サン ホゼ・ミラン ドライブ 108 番・445	
(72) 発明者 エイブラハム・バーレル イスラエル国・99782・ドアール ナ シ ムション・ミシュマール アヤロン・22		(72) 発明者 ローレンス・ケント・ガディ アメリカ合衆国・95118・カリフォルニア 州・サン ホゼ・ラヴェンナ コート・ 5623	
(72) 発明者 エイモス・ブラウン イスラエル国・54000・ジバット スムヘ ル・ハナシ・45		(72) 発明者 フィリップ・ボール・ブヨ アメリカ合衆国・95130・カリフォルニア 州・サン ホゼ・エルムウッド ドライ ブ・4989	

